Safer Internet Lab

Research Report

Online Fraud and Scam Trends Across APAC

Examining the growing role of Artificial Intelligence in the development and proliferation of online scams across the region

Safer Internet Lab

ONLINE FRAUD AND SCAM TRENDS ACROSS APAC

Safer Internet Lab

A Research Report by Safer Internet Lab

TABLE OF CONTENTS

Introduction

Adinova Fauri, Futy Ichiradinda

1

Introduction

Adinova Fauri[1], Futy Ichiradinda[2]

ONLINE FRAUD AND SCAMS IN ASIA PACIFIC

The Asia-Pacific region has become the fastest-growing digital economy in the world, with over 1.8 billion mobile subscribers and contributing approximately 23% to global digital services trade in 2023 (GSMA, 2024a; ADB, 2025). Economies such as China, India, Japan, South Korea, and Singapore are leading this progress through major investments in digital infrastructure, e-commerce platforms, and cloud-based services, thereby positioning the region as a digital trade powerhouse (ADB, 2025; Beschorner, et al., 2019).

However, this rapid digital expansion has also introduced significant security and risks. Institutional frameworks have struggled to keep pace with the acceleration of digital adoption, exposing gaps in cybersecurity preparedness, data governance, and cross-border regulatory coordination. A particularly urgent concern is the sharp rise in online fraud and generative AI-enabled scams such as deepfakes, which one of the growing forms of cybercrime across the region (ADB, 2023; World Economic Forum, 2025). Malicious actors are increasingly deploying generative AI tools, such as deepfakes, voice cloning, and synthetic identity fabrication, to evade conventional fraud detection systems and conduct highly personalized, cross-platform attacks.

These challenges are further compounded at the regional level, which different legal regimes and limited regional enforcement capacity hinder efforts to address online fraud and scams. While some Asia-Pacific nations have begun rolling out cybersecurity frameworks and online fraud and scams, most remain reactive, fragmented, domestic oriented, and sector-specific. As generative AI accelerates the sophistication of cyber threats, national policies alone are proving insufficient to contain risks that are inherently complex and cross-border in nature.

The nature of online scams sometimes varies significantly across Asia-Pacific countries, shaped by national contexts. In Southeast Asia, such as Indonesia, Philippines, Thailand, and Vietnam, the rapid post-pandemic adoption of digital services—such as digital payments, e-commerce, and social media platforms—has coincided with rising public vulnerability to scams, including those powered by artificial intelligence (AI). Conversely, more digitally advanced economies such as Singapore, South Korea, and Australia are increasingly targeted by sophisticated online fraud schemes that exploit economic affluence and high public trust in institutions. In addition, the region also face mounting difficulties in addressing cross-border scams amid fragmented transnational enforcement.

Despite varying levels of digital readiness across the region, a shared concern persists that the online scams and deepfakes has outpaced current regulatory frameworks and institutional coordination. Most countries in Asia Pacific report similar challenges, outdated, sectoral legal frameworks, disjointed regulatory authorities, and insufficient AI governance mechanisms that fail to reflect the latest technological developments. Without strengthened institutional coordination and cross-border collaboration, national capacities will continue to lag behind the evolving landscape of digital threats. Addressing online scams and deepfakes cannot be tackled through national efforts alone. The cross-border nature of many scams and uneven regulatory capacities across jurisdictions demand more adaptive regional cooperation. Initial steps—such as the formation of ASEAN working groups on scams

[1] Researcher, Department of Economics, CSIS Indonesia
[2] Project Research Assistant, Safer Internet Lab, CSIS Indonesia

and AI governance—are commendable but remain insufficient for operational needs like data exchange and minimum digital security standards.

This report is developed in response to the growing need to understand the dynamics of online fraud and AI-driven scams, across Asia Pacific through a case study approach. Each chapter examines the current trend of online fraud and scams, national regulatory landscape, key actors, and challenges in addressing fraud and scams. Through this report, the study of ten national contexts is presented not only to map specific risks and responses, but also to provide a foundation for developing more adaptive, evidence-based, and context-sensitive policy collaboration at the regional level. We believe that this report will help address existing gaps by identifying institutional and policy shortcomings across regions, highlighting promising practices, and outlining key action plans to combat online fraud and scams.

ONLINE FRAUD AND SCAMS: FRAMEWORK AND IMPLICATIONS

Asia-Pacific is among the most cyber-attacked regions in the world. The rapid growth of digital technology with the absence of robust digital governance has created new risks and vulnerabilities, particularly in relation to online fraud and scams. While most countries in the region have established their own digital governance frameworks, these are often reactive, weakly enforced, and domestic oriented. As a results, the region's ability to effectively combat online fraud and scams remains limited. The GSMA's (2024b) Digital Nations Index offers a valuable conceptual framework to assess digital system maturity through five core pillars: infrastructure, institutions, innovation, security, and inclusion.

Table 1.1 Five Core Pillars of Digital Maturity Framework

Countries	Infrastructure	Innovation	Data Governance	Security	People
Indonesia					
Thailand					
Vietnam					
Philippines					
Singapore					
South Korea					
Australia					
Japan					
India					

Source: GSMA Digital Nations Index (2024b)

0-24 25-49 50-75 76-100

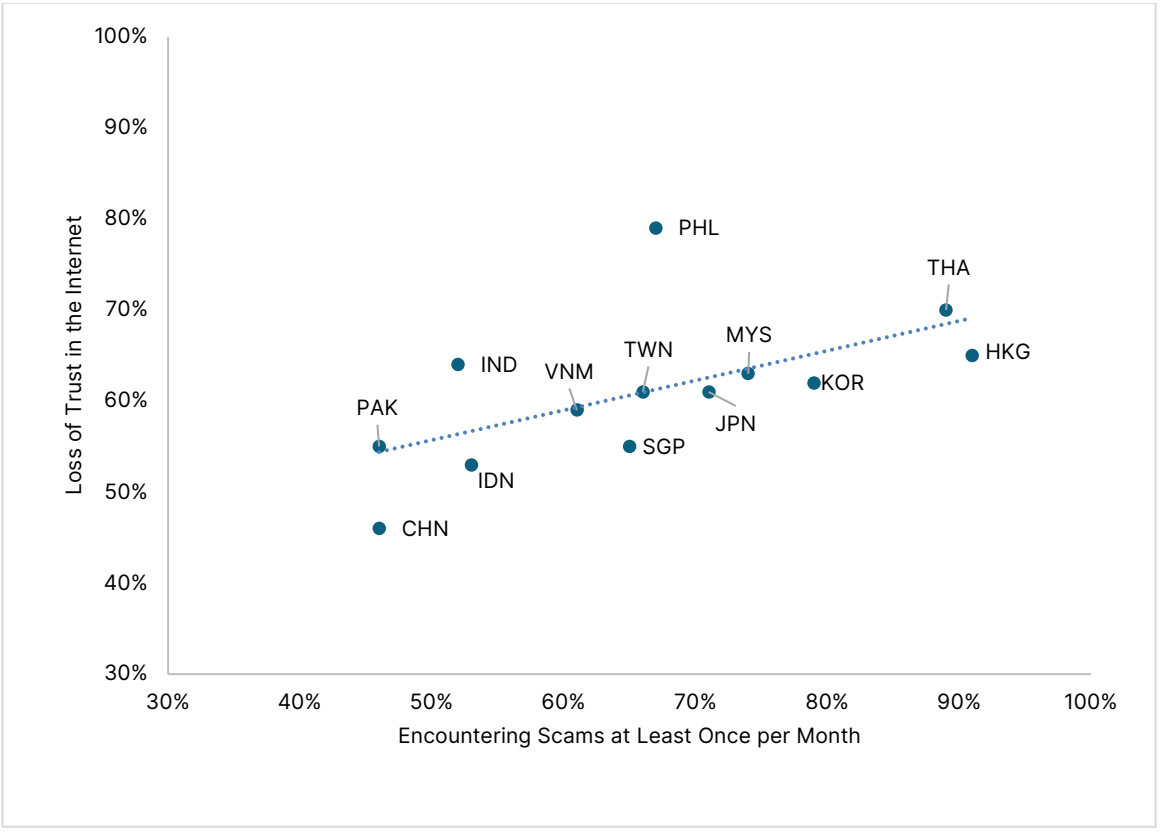
A central challenge in the region is the uneven progress in the digital development and governance. Countries like Singapore, Australia, and South Korea lead across all five pillars, reflecting advanced digital ecosystems supported by strong regulatory and technological foundations. However, even in these more mature contexts, digital fraud and scams remain persistent threats, driven by gaps in digital literacy and a lag between technological advancement and corresponding governance mechanisms.

Meanwhile, other countries in the region continue to struggle with infrastructure gaps, low digital literacy, and weak cybersecurity systems (ADB, 2025; UNESCAP, 2024). In particular, suboptimal data

governance and inadequate cybersecurity measures remain major vulnerabilities, often exploited as gateways for cyberthreats, including online fraud and scams. To address these issues, the five pillars outlined in the GSMA framework should serve as strategic priorities in strengthening regional responses. However, domestic reforms alone are insufficient. Given the borderless nature of online scams, regional and international cooperation is essential. Efforts should focus on harmonizing anti-scam standards, exchanging best policy practices, and developing coordinated mechanisms to enhance collective resilience against cybercrime.

Meanwhile, efforts to strengthen anti-scam measures across the region are becoming increasingly urgent, given the potentially severe socio-economic consequences. The financial losses from fraud and scams in Asia are estimated to reach US$688 billion in 2024 (GASA, 2024). Beyond direct economic impact, these crimes also have broader implications, particularly eroding digital trust. As we can see from the Figure below, there is a clear correlation between the frequency of scam encounters and declining trust in the internet. This erosion of public trust can further undermine the digital economy, diminishing its long-term growth potential.

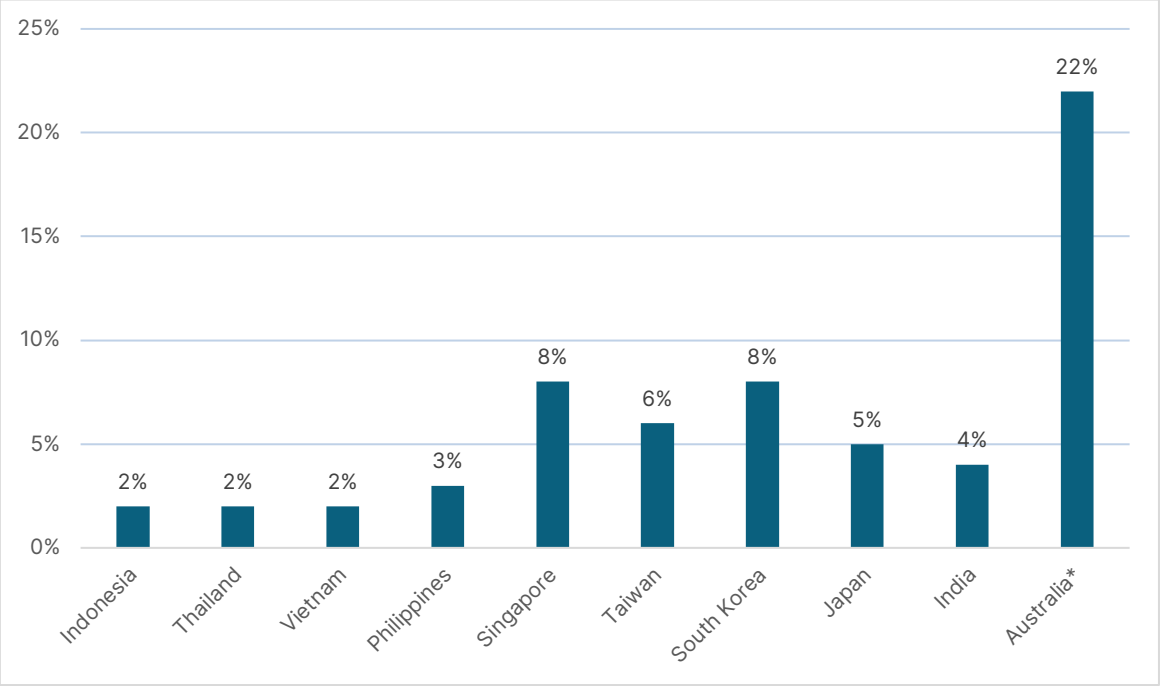
Figure 1.1 Relationship Between Scam Encounters and Loss of Trust in the Internet



Source: GASA (2024)

The threat to digital trust is likely to deepen if governments, private sectors, and other key stakeholders fail to step up their efforts in combating fraud and scams. One of the most pressing issues is the low rate of fund recovery. Currently, there is no effective domestic mechanism in the region for recovering the loss funds for the victims of online fraud. In countries like Vietnam, Indonesia, and Thailand, only around 2% of lost funds are recovered, while India and Japan fare slightly better, although still under 5%. Australia, with a recovery rate of 22%, performs relatively well but remains far below what is needed.

Figure 1.2 Percentage of People Who Successfully Recovered Their Lost Fund



Source: GASA (2024)

Note: While data from other countries was sourced from the 2024 GASA's report, Australia's data is based on the 2023 GASA's report.

Despite these challenges, a number of countries in Asia-Pacific have initiated programs and institutions to counter online scams (see Table below). Indonesia, Singapore, and Thailand have even established anti-scams centers or launched joint initiatives involving multiple stakeholders. However, domestic policy frameworks alone sometimes are not enough. Multistakeholder collaboration, among governments, the private sector, and civil society, is critical, both at the national and regional levels. Such collaboration is needed to build more resilient and trustworthy digital ecosystems across the Asia-Pacific.

Table 1.2 Institutional Mapping for Online Fraud Response

Country	Authority / Institution	Key Regulations / Strategy & Initiatives
Australia	• Australian Competition and Consumer Commission (ACCC) • Australian Federal Police (AFP) • Australian Signals Directorate (ASD) • Australian Communications and Media Authority (ACMA)	• Scamwatch operations; consumer protection and scam reporting • Cybercrime enforcement and international investigation coordination • Cyber threat intelligence, monitoring, and digital infrastructure protection • Regulation of digital platforms, spam, and scam communication channels
India	• CERT-In (Indian Computer Emergency Response Team) • Ministry of Electronics and IT • Reserve Bank of India • Securities and Exchange Board	• Amendment of Information Technology Act • RBI Guidelines on Financial Fraud • AI-powered scam tracking system • AI Ethics Guidelines • Consumer Protection Laws

Indonesia	• Financial Services Authority (OJK) - *Satgas PASTI* • Ministry of Communication and Digital (Komdigi) • National Cyber and Crypto Agency (BSSN) • Bank Indonesia • National Police	• Electronic Information and Transaction Law • Indonesia Anti-Scams Center (IASC) • AI Ethics Framework • Regulation on Electronic Systems • Presidential Regulation and BSSN Regulations on Cybersecurity Management
Japan	• Digital Agency • National Police Agency • Financial Services Agency (FSA) • Consumer Affairs Agency	• National digital infrastructure coordination; e-government service integration • Cybercrime prevention and enforcement; scam investigation via social media platforms • Regulation of financial platforms and monitoring of fraudulent investment schemes • Public awareness campaigns; response to impersonation and fake advertisement scams
Philippines	• Bangko Sentral ng Pilipinas (BSP) • Department of Information and Communications Technology (DICT) • National Telecommunications Commission (NTC) • Department of Justice (DOJ)	• National Financial Inclusion Strategy; scam risk monitoring and consumer protection policies • Oversees digital transformation, cybersecurity, and AI governance planning • Regulation of SMS/telecom-based scams; SIM Registration Act enforcement • Legal prosecution of cybercrime under Cybercrime Prevention Act
Singapore	• Singapore Police Force (SPF) • Ministry of Communications and Information (MCI) • Infocomm Media Development Authority (IMDA) • Monetary Authority of Singapore (MAS)	• National Anti-Scam Command (ASC); centralized scam response coordination • Oversight of digital media, trust and safety, and public communications • Co-regulates with MCI; responsible for digital trust, online safety regulations • Financial scam detection, regulation of digital payment systems
South Korea	• Ministry of Science and ICT (MSIT) • Korea Internet & Security Agency (KISA) • Financial Services Commission (FSC) • Korean National Police Agency	• National Cybersecurity Strategy • AI governance policies • Law enforcement cooperation frameworks • Gaps in regulation on AI scams
Taiwan	• Ministry of Digital Affairs (MDA) • Financial Supervisory Commission (FSC) • National Police Agency	• Ethical guidelines on AI use • Cross-sector initiatives to combat online scams • Regulatory gaps especially in crypto oversight
Thailand	• Ministry of Digital Economy and Society (MDES) • Bank of Thailand • Royal Thai Police • Thai Bankers' Association	• National Cybersecurity Strategy • Multi-factor Authentication Regulation; Industry-wide Anti-Fraud Guideline • Cybercrime Enforcement Protocols; Cross-border cooperation with China on scam suppression • Sectoral Fraud Monitoring Mechanism under the National Anti-Scam Framework

Vietnam	• Ministry of Public Security (MPS) • Ministry of Information and Communications (MIC) • Ministry of Science and Technology • State Bank of Vietnam	• Oversight of cybercrime enforcement; AI-related risk monitoring under national security law • Responsible for digital transformation, AI policy coordination, and public communication • Lead agency for AI Strategy and R&D policy • Financial cybersecurity and digital fraud surveillance

CHAPTER SYNOPSIS

The rapid growth of digital technology in the Asia-Pacific region brings new challenges that demand evolving regulatory frameworks. While these advancements expand digital economic opportunities, they also fuel the rise of online scams that exploit gaps in infrastructure, enforcement, and public awareness. Emerging technologies like Artificial Intelligence (AI) have added new dimensions to these risks, offering efficiencies while also enabling deceptive practices like deepfakes and fraudulent content. This report explores how countries across the region are responding to these threats, highlighting scam trends, policy measures, and recommendations to support more effective national and regional responses.

Australia has become a key target for online scams due to its high-income status, extensive digital adoption, and appeal to AI-enabled fraud. In 2024, government platforms recorded 494,732 scam incidents, with losses reaching AU$2.03 billion. Notably, deepfake videos of public figures were used to promote a fake investment platform. The government responded by establishing the National Anti-Scam Centre (NASC), enacting the Scams Prevention Framework Bill 2025, and stronger collaboration with industry. While Australia's response mechanisms are relatively advanced, challenges persist in protecting vulnerable groups and addressing cross-border scams amid gaps in transnational enforcement.

India with its rapidly growing digital economy and an internet user exceeding 800 million, has become a major target for online fraud and AI-driven scams. According to The Logical Indian report, loses from such scams are expected to exceed ₹20,000 crore in a single year. In response, the government has amended the Information Technology Act to strengthen cybersecurity and The Reserve Bank of India has issued several directives to combat financial fraud. Despite these measures, key challenges remain, including the Act's limited scope in addressing AI-driven scams, lack of effective cross-border enforcement mechanisms, and uneven digital literacy. These issues highlight the need for a more proactive and comprehensive approach to fraud prevention.

Indonesia is increasingly vulnerable to online scams, including those enabled by AI, with high-profile cases such as a deepfake video of President Prabowo deceiving over 100 victims across 20 provinces. In response, the government has enacted the Electronic Information and Transactions Law, the Data Protection Law, and established the Indonesia Anti-Scam Centre (IASC). However, fragmented institutional roles, weak enforcement, and low public awareness remain key challenges, underscoring the need for stronger coordination, cross-sector collaboration, and proactive measures.

Japan has seen a rise in online scams, particularly fake investment ads and fraudulent e-commerce sites that exploit strong public trust in authoritative figures and brands. A notable case involving fake investment seminar ads misusing the identity of a prominent entrepreneur led to 188 reports and ¥2 billion in losses. The government has strengthened advertising regulations, launched public campaigns, and expanded interagency and private sector collaboration. However, challenges persist, including an aging population, deep-rooted trust in public figures, and limited cross-border enforcement.

Philippines has experienced a surge in online scams with the rapid shift toward digitalization, particularly in financial services. In 2024, the Bangko Sentral ng Pilipinas (BSP) reported that 59.4 percent of internet users encountered or fell victim to online fraud. To address these threats, the government has enacted key regulations including the Cybercrime Prevention Act, the Anti-Financial Account Scamming Act, and the National Cybersecurity Plan. However, enforcement is hindered by limited resources, weak inter-agency coordination, and the lack of frameworks specifically addressing AI-related risks, alongside ongoing gaps in public education and awareness.

Singapore faces a high volume of online scams, driven by digital connectivity, affluence, and public trust in institutions. In 2024, over 50,000 cases were reported, including a deepfake video of Prime Minister Lawrence Wong and Senior Minister Lee Hsien-Loong used in a fraudulent scheme. The government responded by enhancing cross-agency cooperation, passing the Protection from Scams Bill, and establishing the Anti-Scam Command (ASCom). Challenges remain in addressing cross-border offenders and regulatory gaps, particularly concerning AI-enabled fraud.

South Korea's fight against online fraud is increasingly challenged by AI-driven scams, including cases such as deepfake videos simulating child abductions and celebrity impersonations used in investment fraud. To address these threats, the government established the Whole-of-Government Task Force on Telecommunication Financial Fraud Response and engages in international cooperation through forums like the Global Fraud Summit. However, efforts are hindered by institutional fragmentation, regulatory gaps in data protection, and limited jurisdiction over foreign platforms, leaving responses to large-scale AI-enabled scams fragmented and reactive.

Taiwan has seen a sharp rise in sophisticated online scams amid growing digitalization, particularly on social media and messaging platforms. As surveillance on phone-based fraud has increased, criminals have shifted to using deepfakes, generative AI, and crypto-related scams, with investment scams involving fake celebrity endorsements becoming prevalent. The government has responded by introducing several regulations, including the Fraud Crime Hazard Prevention Act and Anti-Fraud Guidelines 2.0. However, challenges persist, including cross-border operations, weak enforcement, limited consumer protection, and regulatory gaps in areas such as cryptocurrency, AI, and social media-based fraud.

Thailand's growing concern over online fraud, highlighted by an AI-generated scam call targeting Prime Minister Paetongtarn Shinawatra. To address these threats, Thailand has launched key interventions, including the Anti-Online Scam Operation Center (AOC 1441) under the Ministry of Digital Economy and Society (MDES), and strengthened cross-border cooperation through a joint cybercrime crackdown with China near the Myanmar border and participation in ASEAN's working group on AI and online fraud. Nevertheless, challenges persist, particularly due to the lack of enforceable AI regulations.

Vietnam is facing a growing cyber threat from increasingly sophisticated online scams, driven by generative AI and targeting vulnerable groups such as SMEs. One major cross-border case involved a Cambodia-based fraud operation using deepfake impersonations of Vietnamese officials, defrauding over 13,000 victims and resulting in estimated losses of 1 trillion VND. While Vietnam has established a legal foundation through the Cybersecurity Law and the Personal Data Protection Decree, current efforts lack targeted measures to address AI-enabled fraud. Key challenges include fragmented coordination, low public awareness, and the transnational nature of scams, highlighting the need for a coordinated multi-stakeholder response.

REFERENCES

ADB. (2023). *E-commerce Evolution in Asia and the Pacific: Opportunities and Challenges.* Asian Development Bank. From https://www.adb.org/publications/e-commerce-evolution-asia-pacific-opportunities-challenges

ADB. (2025). *The Role and Future of Digital Economy Agreements in Developing Asia and the Pacific*. Manila: Asian Development Bank. From https://www.adb.org/publications/digital-economy-agreements-asia-pacific

Beschorner, N., Bartley Johns, M., Guermazi, B., Treadwell, J. L., Prakosa, P. W., Abdul Karim, N. A., . . . Girot, C. A. (2019). *The Digital Economy in Southeast Asia : Strengthening the Foundations for Future Growth.* Washington, D.C.: World Bank Group. From http://documents.worldbank.org/curated/en/328941558708267736

GASA. (2023). *The State of Scams in Australia.* Global Anti-Scam Alliance. From https://www.gasa.org/research

GASA. (2024). *Asia Scam Report 2024.* Global Anti-Scam Alliance. From https://www.gasa.org/research

GSMA. (2024a). *The Mobile Economy Asia Pacific 2024.* GSMA. From https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/asiapacific/

GSMA. (2024b). *Digital Nations in Asia Pacific: Preserving digital trust.* GSMA. From https://www.gsma.com/about-us/regions/asia-pacific/gsma_resources/digital-nations-2024/

UNESCAP. (2024). *Asia-Pacific Digital Transformation Report 2024.* United Nations Economic and Social Commission for Asia and the Pacific. From https://repository.unescap.org/server/api/core/bitstreams/95991d3a-149c-4dcf-ae2b-9250c5c04614/content

World Economic Forum. (2025). *Global Cybersecurity Outlook 2025*. World Economic Forum. From https://www.weforum.org/publications/global-cybersecurity-outlook-2025/

Online Fraud and Scams in Australia

Billy Esratian

2

Online Fraud and Scams in Australia

Billy Esratian[3]

INTRODUCTION

Australia is deemed to be a lucrative market for online scams. In 2022-2023, the Australian Bureau of Statistics estimates that Australia's real net national disposable income per capita reached AU$71.774. Such a relatively wealthy society serves as one of the motivating factors for online scammers to target Australians in their operations. The lure for online scammers to target Australia is also induced by Australia's highly digitized society, with internet users' proportion in 2023 reaching up to 97,1% of its population (International Telecommunication Union, 2024). Such a high dependence into the digital online activities concurrently allows the scammers to have more entry points to reach the Australian society.

Current trends of online scams in Australia witness an accelerated diversification method to primarily obtain and exploit the target's personally identifiable information (PII) for illegal gains. This diversification effort is facilitated via the use of generative AI by the scammers. Harnessing the collaborative ecosystem and multilayered approaches between the government, private sectors, law enforcement and the public in general, efforts to address this AI-powered online scams are underway. While the improvement of the situation is evident, several challenges to respond the ever-advancing scams operations still need to be addressed.

PATTERNS AND TRENDS

As a disclaimer, there are several reporting platforms available for Australians to report scams, namely through Scamwatch, ReportCyber, IDCARE, Australian Financial Crimes Exchange (AFCX), and Australian Securities and Investment Commission (ASIC). In 2024 alone, these collective reporting mechanisms garnered a total of 494.732 reports, with losses accumulating up to AU$2.03 billion (National Anti-Scam Centre of Australia [NASC], 2025). Given that uniformed methodology and data consistency are essential to observe patterns and trends, for the purpose of the *Patterns and Trends Section*, the Scamwatch data as one of the derivatives of all said reporting mechanisms will be solely used as the baseline study for the subsequent discussions. This is because Scamwatch wields a relatively more consistent and detailed extent of data compared to the other reporting mechanisms.

In the aggregate, throughout the space of 2023 to February 2025, Scamwatch which is run by the NASC, received 587.744 reports of scams with losses totalling AU$ 861,022,129.19. Delving into the data, three types of scams contribute the most to the losses, namely investment, dating and romance, and phishing scams. Despite the differences of conduct in these three types of scams, one commonality remains, for in each type of these scams, the use of generative AI was evident to a certain extent, and served as an early precaution of the trends that may lurk ahead for one's personal safety.

Investment Scams

Investment scams dominated the typology of reported scams in Australia, with losses amassing at AU$ 519,126,719.13 from 17,382 reported investment-related scams (NASC, 2025). The median number of losses is estimated to be around AU$8500-AU$10,000. Two types of scams are prevalent in this corpus:

[3] Ph.D. in Law Candidate, The Australian National University

a) the use of convincing marketing and advanced technology, with promises of big payouts and little losses, done with pressure tactics; and b) gambling made in the form of investment scam, through computer prediction software, betting syndicates, and sports investment (NASC, 2025).

As a modus operandi, the use of AI was identified as an emerging trend in investment scams, particularly, by way of AI trading platform scams (RMIT, 2023). To this end, the scammers claim that the trading platform is harnessed using AI software and other emerging technology, such as the quantum computing, to maximize the returns, and therefore would compensate the lack of expertise from the investors' side. In luring these inexperienced investors, the scammers use deepfake prominent figures' endorsements to initiate the online trading. One notable example in Australia was the use of AI-engineered doctored videos of Elon Musk and Chris Hemsworth that promoted the Quantum AI, a fake investment trading platform (ASIC, 2024). In 2023 alone, it was estimated that the losses from AI trading platform scams amounted to nearly AU$20 million, with more than 600 reports to NASC's Scamwatch noting the use of common methodology in AI trading platform scams (RMIT, 2023).

Dating and Romance Scams

Dating and romance scams contribute second to the cumulative losses due to scams in Australia (Scamwatch, 2025). Out of 7.403 reports, the losses peaked at AU$63,709,074.97 with the median of AU$1,792.50. Dating and romance scams manifest through a perceived real and genuine relationship offered by the scammers to the victim. (Cross & Layt, 2022) Once the trust is gained, manipulations to give the scammers money, gift, or personal data will then proceed as the subsequent step. The avenues to practice this scam are diverse, encompassing social media, gaming and dating apps or websites, and direct text and email. To achieve the objective, the scammers often disguise themselves in a fake identity, including famous person. (Cross & Layt, 2022).

The Emerging use of AI is particularly apparent to sustain the romance scams in two models. First, through the deepfake technology to help create a new identity. Second, through the use of Chatbots in the conversation. To this end, two indicants may hint said usage of AI, namely flawless looking photos, and vague and repetitive answers (Australian Banking Association, 2025). In one instance, the use of deepfake AI was noted to be as sophisticated as materializing into a live video call (ABC, 2024). This is particularly concerning because the use of deepfake video begins to erode the traditional predicates to detect romance scammers, such as the scammers' avoidance to meet in person or to have a video call. Another trend observable from Australia is that romance scams may be cross-jurisdictional in its practice. In one instance, some 5000 Australians were targeted as potential victims of the Philippines-based romance scammers. (Australian Federal Police [AFP], 2025).

Phishing Scams

The reported losses caused by phishing scams amounted to AU$54,918,571.93 as accumulated from 218.591 reports, with the median losses reaching up to AU$2,345.60. (Scamwatch, 2025). Phishing scams use impersonation to obtain certain information from the target, usually through the sending of a message that is characterized as if it is coming from a solid entity, claiming the sense of urgency and importance, and inviting the target to open a deceptive website (Desolda et al., 2022). Phishing also often includes a stressor statement to stimulate the recipient to act promptly without proper consideration (Grimes, 2024).

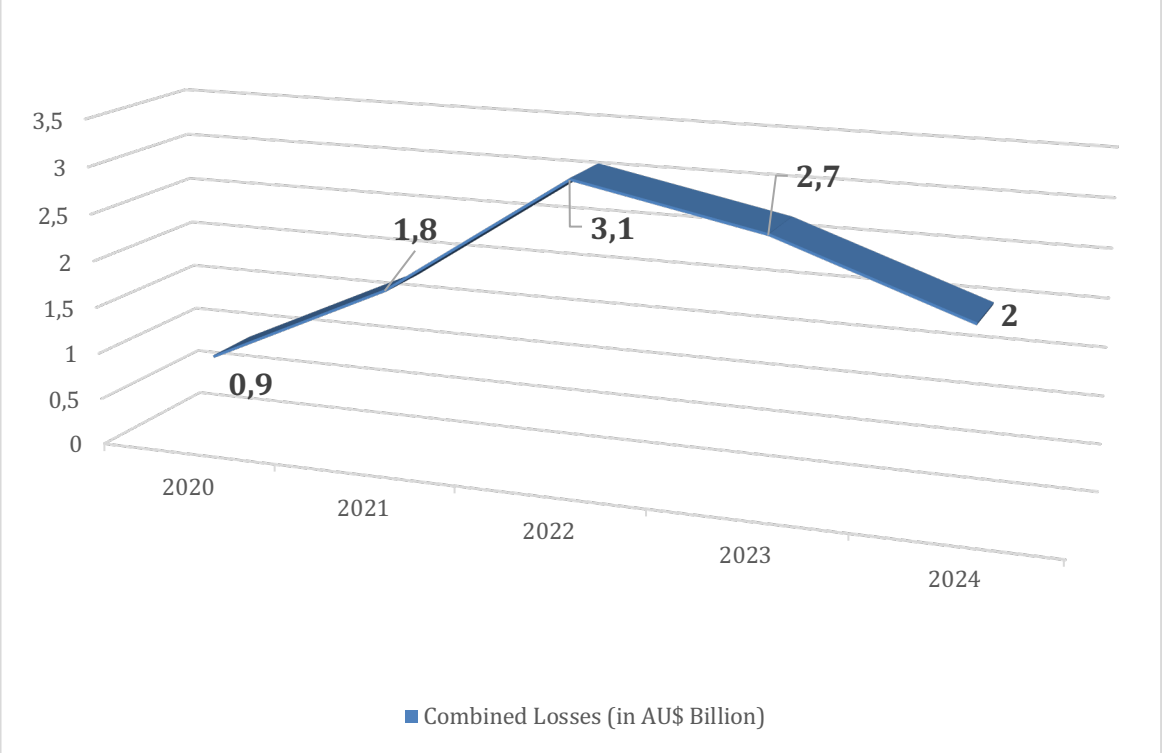
Various models of phishing scam's sophistication can be attributed to the use of generative AI (Das, 2024). Corollary to this trend, the Australian Signals Directorate detected a variant of phishing, called Vishing (video phishing) wherein the scammers use the deepfake technology to interact with the victim

via video conference to dupe the victim into doing what is told by the scammers (ASD, 2024). One example of this vishing technique impersonated the Sunshine Coast Mayor, Rosanna Natoli. In this regard, the scammer(s) created Facebook accounts, engaged in live video calls through facial alteration technology to mimic Rosanna Natoli, and requested bank details via messenger (ABC, 2024). Another notable trend of phishing variants in Australia is the Quishing (Quick Response (QfR) Phishing) which uses the QR code technology to deceive the target into giving a personal information or downloading malware to the target's device (ASD, 2024). One example of this practice involved the scam email impersonating the Australian Taxation Office with QR code directing the target to a fake myGov login page, with the purpose of stealing the target's myGov account details (ASD, 2024). Through the spreading of publicly sourced AI-powered QR generators, the practice of QR Phishing thus needs to also be anticipated. In another observation, the exponential increase of AI-generated phishing text message scams activities may contingent upon certain occasions. In this context, the AFP highlights that sales events, such as the Black Friday and Cyber Monday may be exploited through the impersonation of legitimate retailers and postal and delivery services (AFP, 2024).

IMPLICATIONS OF ONLINE FRAUD AND SCAMS

At the outset, economic losses are one indicant of the negative impact that scams operations brought to the society. To this end, Australia experiences a relatively sizable economic losses from scams where the yearly losses can reach as high as AU$3,1 billion.

Figure 2.1 Combined Losses from Scams in Australia



Source: NASC (2025)

Beyond said economic losses, three societal and legal ramifications further exacerbate the impacts of online scams in Australia, namely the disproportionate victimisation of the elderly generation, the heightened risk on personal safety, and the cross-jurisdictional issue in law enforcement.

The Disproportionate Victimisation of the Elderly Generation

Scams affect different age groups dissimilarly. Focusing on the three types of scams from 2023-February 2025 as previously discussed, Scamwatch data revealed that the elderly generation (age 65 and over) is disproportionately affected by online scams. This group accounts to 68.639 out of 243.376 reported scams, with losses amounted to AU$196,979,844.96 (Scamwatch, 2025). This figure is even higher than the number of reports of the other five age groups (under 18 (690 reports), 18-24 (5.056 reports), 25-34 (13.516 reports), 35-44 (20.368 reports), 44-54 (24.999 reports)) combined (64.629 reported scams) (Scamwatch, 2025).

A variety of explanations may be offered to rationalize why the older generation is more susceptible to online scams. From financial standpoint, this generation is relatively considered to be a suitable target due to their relatively stable finance from the accumulated lifesavings and retirement funds (Button et al, 2024). From medical standpoint, cognitive impairments along with health problems may increase their risk of victimisation (Button et al, 2024). From psychological standpoint, the condition of living alone or lack of social networks may also add as an aggravating factor. (Button et al, 2024). Additionally, misplaced trust and lack of digital literacy may also be invoked as alternative explanations (ANZ, 2024).

While the rationale may diverge, the disproportionate victimisation of online scams targeting the elderly generation in Australia needs to be meticulously and promptly addressed, particularly in the wake of the incorporation of AI in online scam schemes. The data shows that this group of population is the most vulnerable and susceptible to online scams. Tailored policy and intervention need to be the paradigmatic approach to address the predicament, and to prevent further illegal encroachment to this generation's hard-earned savings, tranquillity, and time.

The Heightened Risk on Personal Safety

Recent online scams operations have seen a diversification of entry point to reach the target. Scamwatch data shows that while email (194.501 reports) and text messages (191.227), with a combined number of 385.728 reports, remain the top contact methods for scams throughout 2023-February 2025, social media (38.311) and mobile apps (14.247) contribute considerably as another entry points for scams to the overall reports with 52.558 reports (NASC, 2025).

The use of social media and mobile apps for malicious purposes, such as scams, presents another unsolicited intrusion on the personal space, and therefore safety. In one method, the operation involves the sending of direct messages through LinkedIn, offering work-from-home job offers and in return, requesting the sending of personal data as one of the requirements to accept the job offers (Cross, 2014). In another method, one Sydney bank worker lost AU$157,000 as she was tricked to trade in a cryptocurrency scam by a verified profile account persona that she met on Tinder (9News, 2023).

People use different social media for different purposes, including but not limited to the search for entertainment, professional advancement, and social interaction (Sheldon, 2015). The pervasive spread of online scams in various social media and mobile apps, suppresses the attainment of these varied personal objectives of social media use, and put the users at risk. The responsibility to eradicate the spreading of scams via social media and mobile apps should not lie solely on the users, a significant portion of it should also be rendered to the online platforms. This includes the improvement over the security of the platforms, such as enhancing the reliability of users' verification, and the detection of scam patterns and activities, especially in the wake of AI-powered scams.

The Cross-Jurisdictional Issue in Law Enforcement

Online scams are traditionally considered to be a low to medium risk crime for the perpetrator, primarily because direct encounter between the target and the perpetrator is not necessarily needed. Furthermore, with the usage of AI, the traceability of the crime leading to the direct perpetrator is even more convoluted. This indirect nature of interaction, often involving third parties, such as email and social media platforms, flourishes the cross-jurisdictional proclivity for online scams operations. Particularly in Australia's context, such a cross-jurisdictional scams operations are motivated to target Australia given the comparative wealth that Australia has in the region (The Australian Transaction Reports and Analysis Centre [AUSTRAC], 2024).

AUSTRAC estimates that the scams targeting Australians are substantially organized by offshore criminals, particularly the transnational serious and organised crime groups (AUSTRAC, 2024). In multiple online scams cases targeting the Australians, AUSTRAC's estimate is confirmed. In one cryptocurrency scam case, the culprit launched his operation from the Ukraine by specifically targeting Australians (ABC, 2024). In another case, particularly in romance scam case, some 250 suspects were based in Manila, the Philippines and managed to potentially swindle around 5000 Australians (AFP, 2025). Another account also noted the large-scale scam operations from Cambodia, with around 100,000 people confined in cyber-scam compounds around Cambodia, and are forced to engage in cybercriminal activities, including fraudulent investment, romance and cryptocurrency scams (United Nations Office on Drugs and Crime, 2023).

This cross-jurisdictional operations may present a challenge for Australia's law enforcement to immediately act, for traditionally, law enforcement measure will be led by the law enforcement authority of the respective State where the crime or the criminal is situated. To this end, offshore criminal intelligence sharing is vital. However, several issues may still hamper this effort, including the transactional tendency of intelligence sharing and the lack of mutual trust between offshore law enforcements (Phil et al., 2018). While a notification on possible scams operations to the respective State may be useful to start build the trust and assist said law enforcement measure, a more robust and well-rounded approach may yield a more promising and expedient result. To this end, the institutionalization of both bilateral and regional assistance needs to be considered in order to garner trust and close the looming disparity of law enforcement's capacity between States in the region.

THE ROLE OF KEY STAKEHOLDERS AND STRATEGY IN ADDRESSING ONLINE FRAUD AND SCAMS

Key Stakeholders Mapping in Australia

The Australian Competition and Consumer Commission (ACCC), the Australian Signals Directorate (ASD), and FPD are deemed to be the key actors involved in addressing AI-generated online scams and fraud in Australia. Combined, these three institutions hold the regulatory, administrative, and law enforcement powers requisite to generally prevent and disrupt the spread of online scams in digital platforms. It should also be noted that various other sectorized institutions contribute to the scams control in Australia, including, but not limited to AUSTRAC, ASIC, and the Australian Communications and Media Authority (ACMA).

Table 2.1 Australia's Governmental Initiatives to Address Online Scams

Stakeholder	Statutory Basis	Strategic Role on Online Scams Control
ACCC	Competition and Consumer Act 2010.	Leads the NASC.
ASD	Public Governance, Performance and Accountability Act 2013; and the Intelligence Services Act 2001	Prevent and disrupt offshore cyber-enabled crime; runs the Australian Cyber Security Centre; coordinates with NASC.
FPD	Federal Police Act 1979	Enforcing criminal law; leads the Joint Policing Cybercrime Coordination Centre; coordinates with NASC.
AUSTRAC	Anti-Money Laundering and Counter-Terrorism Financing Act 2006.	Involved when online scams use money laundering technique to transfer and layer the criminal proceeds; coordinates with NASC.
ASIC	Securities and Investments Commission Act 2001; Business Names Registration Act 2011; Corporations Act 2001; Insurance Contracts Act 1984; National Consumer Credit Protection Act 2009; and Financial Accountability Regime Act 2023	Regulate and enforce scams related issues in financial product and services, such as the taking down of investment scam websites, and evaluation over banks' scams prevention mechanism; seek civil penalties and prosecute offenders for the purpose of consumer protection in financial product and services; coordinates with the NASC.
ACMA	Telecommunications Act 1997; Broadcasting Services Act 1992; and Australian Communications and Media Authority Act 2005.	Register and enforce rules to telecommunications sector on scams issues; coordinates with the NASC.

Australian Competition and Consumer Commission

ACCC is tasked with the administering and enforcing of primarily, the Australia's Competition and Consumer Act 2010. In this regard, the ACCC for instance managed to bring a case against Qantas pertaining to the offering and selling of an already cancelled flight tickets which misled the consumers (Federal Court of Australia, 2024). Additionally, since 2023, the ACCC is also mandated to run the Australia's NASC, a collaborating platform involving representatives from the industry and government to disrupt scammers and raise awareness on the issue.

Australian Federal Police

The AFP can be classified within the law enforcement function as it enforces Australia's criminal law, contributes to combat the transnational, serious, and organised crime impacting the Australia's national security, and protect Australia's interest from criminal activity both home and offshore. In relation to online scams, AFP by collaboration with partners both home and abroad, was particularly involved in Operation Aquila which addresses cybercrimes, including identity fraud, Operation Nebulae which brough about the global take down of a phishing-as-a-service platform known as LabHost. and Operation Firestorm, which concentrated to cyber criminals and human trafficking targets in Southeast Asia and Eastern Europe (AFP, 2024).

Australian Signals Directorate

ASD is a government agency which holds the responsibility for foreign signals intelligence and cyber security. In relation to online scams issue, within the ASD, sits the Australian Cyber Security Centre (ACSC), which is tasked as the technical authority of the government on cyber security. In contributing to scams control, ACSC is involved in several related services, including the reporting mechanisms via CyberReport and 1300 CYBER1 (1300 292 371), and the updating of relevant alerts, advisories, and notifications on potential cyber security threats.

Strategy in Addressing Online Fraud and Scams

Three approaches form the hallmark of Australia's current response to online scams, encompassing the regulatory framework, institutional governance, and law enforcement. While generally deemed to having brought about positive outcome, assessment on these approaches is still warranted to detect possible space for improvement.

Regulatory Framework

Scams Prevention Framework Bill 2025

Receiving the Royal Assent on 20 February 2025, the Bill came into force on 21 February 2025. Under this Bill, providers are obliged to take steps to comply with the standardized principles to protect Australians from scams. These principles encapsulate the governance arrangements relating to scams and the prevention, report, disruption, and response to scams. This Bill also tasks the government to create an enforceable Scams Prevention Framework Code containing more detailed provisions on the principles, including sector-specific requirements for the service providers (Scams Prevention Framework Bill 2025).

While the Bill is still relatively new, some concerns are levied upon this Bill, such as the legislation process which was deemed to not substantially take into account the inputs from public consultation by the Senate Standing Committees on Economics, the need for a 12-month transitional period from the industry, clarification for '*actionable scams intelligence,*' and the risk of misconduct by vexatious or speculative litigants within the system (GT Law, 2025).

Institutional Governance

National Anti-Scam Centre

The Australian Government in July 2023 invested AU$58 million for the establishment and the running of the NASC. Sitting within and led by the ACCC, the NASC orchestrates the coordinated measures by various entities, both from within the government and representatives from the industry in controlling scams. One of the products of the NASC is a reporting mechanism named the Scamwatch, which data is essential to the overall reading of scams pattern in Australia. Additionally, the NASC establishes various fusion cells, time-constraint task forces designed to address specific scams. Apart from the collaboration and disruption function, NASC also holds outreach function to raise the awareness on scams. However, the disproportionate victimisation of the elderly generation indicates the need for Anti-Scam Centre to diversify its method of public awareness campaign and its presence within the society.

Safe-Scam Accord

Launched in 2023 and applicable to all members of the Australian Banking Association and the Customer Owned Banking Association, this private sector-led initiative provides common thresholds for Australian banks in addressing scams and specific timeline for its fulfilment. These thresholds cover the disrupt, detect, and respond strategies, whereby some of the determined outputs of these strategies include, the name checking technology to mitigate the potential scam when the name does not match, the use of biometric check for new individual customer to verify the customer's identity, scams intelligence sharing involving the AFCX, and the limiting of payments to high-risk channels such as some crypto currency platforms (Australia Banking Association, 2023).

However, the advancement of scams tactics and methods may have surpassed the security thresholds set within this Accord. In one scam case, a person was duped to disclose two six-digit passcodes to a scammer who was able to contact him through a 1300 phone number. Such a 1300 phone number drew

a similarity with the phone number used multiple times in legitimate messages by the HSBC Bank Australia Limited. Following the disclosure of passcodes, an unauthorized transaction was made by the scammer. The issue was brought to AFCA. In its Determination, AFCA holds that the complainant is entitled for compensation from the bank in this sophisticated bank impersonation scam for the involuntary disclosure of his passcodes and for the bank's conduct in mishandling said scam issue (AFCA, 2024).

Law Enforcement

Joint Policing Cybercrime Coordination Centre (JPC3)

Under the coordination of the AFP, JPC3, which was launched in March 2022, coordinates Australia's policing response to high volume cybercrime. To do so, JPC3 collaborates with representatives from relevant government agencies, banking and financial sectors, and other key stakeholders, including international law enforcement agencies. One example of this collaborative ecosystem was Operation WICKHAM with the United States Secret Service that pursued an attempt to launder US$100 million stolen from victims who invested in a global investment scam (AFP, 2024). In another case, through Operation Guardian, JPC3 collaborated with various private entities to minimise the misuse of PII due to the data breaches in Medibank, Optus, MyDeal, Latitude; and Go-Anywhere.

Nevertheless, while serves as a collaborative model, JPC3 prioritises high volume cybercrimes. Such a threshold provides a room for JPC3 to determine which case to pursue based on its internal discretion and assessment. In online scams cases, the perpetrators are not always associated with sophisticated syndicates. Many acts on their own and separately, may not necessarily be deemed to have caused systematic losses. This is a trend that JPC3 does not necessarily cover.

BEST PRACTICES AND POLICY RECOMMENDATION

With economic losses of reported scams continue to decrease from 2023 to February 2025, Australia's approach to online scams indicates a degree of efficacy. This progress can be attributed to three traits of Australia's effort to control online scams, namely the whole-of-nation approach, the multilayered reporting mechanism, and the data-centric movement.

Best Practices

Whole-of-Nation Approach

Scams control becomes a priority of all fabrics of the State, not only the government. As discussed in the previous section, private entities-led initiative also colours Australia's overall effort to combat online scams, such as through the Safe-Scam Accord. In the government-led initiative, such as the NASC, representatives from the industries are also given with a say in helping to identify the surging trend of online scams. This indicates the blurring of the traditional work division, making the combating of online scams an interest of all, which simultaneously contributes to the raising of awareness among the population, wherein Commonwealth Bank research finds that 60% of Australians are now more concerned about online scams than a year ago (Commonwealth Bank, 2025).

Multilayered Reporting Mechanisms

There are multiple reporting mechanisms available for scams cases, from Scamwatch, ReportCyber, IDCARE, to others. While some of these reporting mechanisms may eclipse each other, one key takeaway that may serve as the best practice is that these multiple reporting avenues provide Australians with more access and alternatives to respond to scams. These multiple reporting mechanisms also allow the Australian Government

to amass and wield a substantially comprehensive set of data pertaining to scams from the submitted reports, from the prevalent modus operandi to the contour of the most impacted group in the society.

Data-Centric Movement

The accumulated data from Australia's scams reporting mechanisms equip the government to periodically evaluate its effectiveness in addressing scams issue. The evaluation allows the government to recalibrate upon the changing landscape of online scams, as indicated by the passing of the Scams Prevention Framework Bill 2025. Additionally, to a certain extent, a significant portion of said data is also widely disseminated to the public, allowing anti-scam initiatives driven by the community to flourish.

Policy Recommendations

Australia's robust intervention, particularly since 2023, has done reasonably well in attempting to curb the losses from online scams. To improve this, several additional measures can still be considered to size up the government's response's efficacy against the AI-powered online scams.

Enhanced Collaboration with Private Sector

Private sector contributes significantly to the developing of generative AI. Some of the latest breakthroughs in various disciplines, such as the use of AI for flood forecasting, detecting wildfires, to teaching personalization, can be attributed to private entities-led AI development initiatives. Through their inventive nature, these private entities should also be involved in the overall effort to combat scams. Additionally, to entities serving as online platforms, their collaboration is indispensable to help detect and supress the spread of scams.

Development of AI for Law Enforcement Purpose

Contrary to its malicious use, AI can also be used to counter crimes. The AFP in collaboration with Monash University for instance established the AiLECS (the Artificial Intelligence for Law Enforcement and Community Safety) Lab to develop AI for automated detection and triage of child sexual exploitation material (AFP, 2022). Initiatives to develop the use of AI for law enforcement purpose like AiLECS should be multiplied by the relevant stakeholders and expanded to address other forms of crimes, including scams. Collaborating with and maximising the resources in the academics and industries, law enforcement apparatus can be benefited in developing AI-driven counter measures to disrupt the use of deepfake, chatbots technologies often used by scammers.

Elderly Generation Intervention

The disproportionate victimisation of the elderly generation indicates a concerning knowledge disparity between age groups on the current scams control response. A more robust intervention is therefore needed. Identification upon the avenues which the elderly generation engage the most should be a priority and the baseline for the NASC to intervene. This intervention should encompass a nation-wide awareness campaign for a family-oriented protection mechanism for the elders. Members of the family should be encouraged to proactively check on the elders on the digital activities that they are engaging with. A fusion cell specifically tasked to decrease the proportion of the elderly generation as victims of online scam can be considered.

Regional Arrangement

From multiple examples as previously discussed, scammers targeting Australians can be based offshore, particularly from across Southeast Asia. Regional policy framework to tackle these offshore criminals needs to be further materialized. One viable avenue to do so is by maximising the Plan of Action to Implement the Association of Southeast Asian Nations (ASEAN)-Australia Comprehensive Strategic Partnership (2025-2029). The detailing of this regional arrangement should address multiple contemporary issues within the spheres of cyber security and transnational crime, such as, common definition and threshold on the malicious use of AI,

the mechanism for joint law enforcement operations, and joint research initiative to develop generative AI to counter AI-powered scams.

Bilateral Support

There exists a considerable capacity gap between states in handling online scams. With the relatively advanced mechanism that Australia has established in eradicating scams, technical assistance to the government of other states, particularly where scammers predominantly operate, should take place. This can be achieved through the secondment of relevant Australian personnels in the respective state for knowledge sharing, and where possible the institutionalization of *ad-hoc* joint operations or a more permanent criminal intelligence exchange mechanism.

REFERENCES

Articles and Reports

Association of Southeast Asian Nations, Plan of Action to Implement the ASEAN-Australia Comprehensive Strategic Partnership (2025-2029) (2024).

Attorney-General's Department, National Plan to Combat Cybercrime 2022 (2022).

Australia and New Zealand Banking Group Limited, Cyber Security for Seniors, available at https://www.anz.com.au/security/protect-your-family/seniors-cyber-security/#:~:text=This%20could%20be%20due%20to,the%20signs%20of%20a%20scam, last accessed on 28 March 2025.

Australian Banking Association, Protect Your Heart and Bank Account from Romance Scams, available at https://www.ausbanking.org.au/protect-your-heart-and-bank-account-from-romance-scams/#:~:text=ABA%20CEO%20Anna%20Bligh%20said,paradise%2C%E2%80%9D%20Ms%20Bligh%20said, last accessed on 28 March 2025.

Australian Banking Association, Scam-Safe Accord, available at, https://www.ausbanking.org.au/new-scam-safe-accord/, last accessed on 28 March 2025.

Australian Broadcasting Corporation, Australians Targeted for Cryptocurrency Scams by Overseas Call Centres because They are 'Easy Prey', Former Worker Says, available at https://www.abc.net.au/news/2024-10-07/scammers-are-targeting-australians-in-offshore-call-centres/104406170, last accessed on 28 March 2025.

Australian Broadcasting Corporation, Authorities Warn AI, Deepfake Technology in Romance Scams Costing WA Victims Thousands, available at https://www.abc.net.au/news/2024-08-28/deepfake-ai-used-in-wa-romance-scams/104279902, last accessed on 28 March 2025.

Australian Broadcasting Corporation, Scammers Use Artificial Intelligence to Impersonate Sunshine Coast Mayor as Experts Warn of Video Call Cybercrime Tactic, available at https://www.abc.net.au/news/2024-05-02/scammers-ai-impersonate-sunshine-coast-mayor-video-call/103794690, last accessed on 28 March 2025.

Australian Bureau of Statistics, Australian National Accounts: National Income, Expenditure and Product (https://www.abs.gov.au/statistics/economy/national-accounts/australian-national-accounts-national-income-expenditure-and-product/mar-2024) (2024).

Australian Competition and Consumer Commission, Report of the National Anti-Scam Centre on scams data and activity 2024 (2025).

Australian Cyber Security Centre, Types of Scams, available at https://www.cyber.gov.au/learn-basics/watch-out-threats/types-scams, last accessed on 28 March 2025.

Australian Federal Police, AFP Reveals 'Rom-Con' Script Used to Scam Victims on Dating Apps, available at https://www.afp.gov.au/news-centre/media-release/afp-reveals-rom-con-script-used-scam-victims-dating-apps#:~:text=The%20script%20was%20uncovered%20during,with%20on%20online%20dating%20apps, last accessed on 28 March 2025.

Australian Federal Police, AFP Submission to the Parliamentary Joint Committee on Law Enforcement: Inquiry into the Capability of Law Enforcement to Respond to Cybercrime (2024).

Australian Federal Police, AFP Takes the Fight to Cybercriminals in 2024, available at https://www.afp.gov.au/news-centre/media-release/afp-takes-fight-cybercriminals-2024, last accessed on 28 March 2025.

Australian Federal Police, Australian Federal Police Annual Report 2021-22 (2022).

Australian Federal Police, Beware of Black Friday and Cyber Monday Scams, available at https://www.afp.gov.au/news-centre/media-release/beware-black-friday-and-cyber-monday-scams, last accessed on 28 March 2025.

Australian Securities and Investments Commission, ASIC Enforcement and Regulatory Update July to December 2024 (2025).

Australian Signals Directorate, Annual Cyber Threat Report 2023–2024 (2024).

Australian Transaction Reports and Analysis Centre, The Money Laundering in Australia National Risk Assessment 2024 (2024).

Button, Mark, Vasileios Karagiannopoulos, Julak Lee, Joon Bae Suh, and Jeyong Jung. 2024. "Preventing Fraud Victimisation against Older Adults: Towards a Holistic Model for Protection." *International Journal of Law, Crime and Justice* 77 (June):100672. https://doi.org/10.1016/j.ijlcj.2024.100672.

Commonwealth Bank, Less Than 10% of Aussies Would Discuss Their Scam Experience with Family, available at https://www.commbank.com.au/articles/newsroom/2025/03/scams-talk-to-a-loved-one.html, last accessed on 29 March 2025.

Cross, C., & Layt, R. (2021). "I Suspect That the Pictures Are Stolen": Romance Fraud, Identity Crime, and Responding to Suspicions of Inauthentic Identities. Social Science Computer Review, 40(4), 955-973. https://doi.org/10.1177/0894439321999311 (Original work published 2022).

Cross, Michael. 2014. *Social Media Security: Leveraging Social Networking While Mitigating Risk*. Waltham, MA: Syngress.

Das, Ravindra. 2024. *Generative AI: Phishing and Cybersecurity Metrics*. 1st ed. Cyber Shorts Series. Boca Raton: Taylor & Francis Group.

Desolda, Giuseppe, Lauren S. Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. 2022. "Human Factors in Phishing Attacks: A Systematic Literature Review." *ACM Computing Surveys* 54 (8): 1–35. https://doi.org/10.1145/3469886.

Gilbert + Tobin Law, The Scams Prevention Framework Legislation Passes Parliament: Time to Get Your House in Order, available at https://www.gtlaw.com.au/insights/the-scams-prevention-framework-legislation-passes-parliament-time-to-get-your-house-in-order2, last accessed on 28 March 2025.

Grimes, Roger A. 2024. *Fighting Phishing: Everything You Can Do to Fight Social Engineering and Phishing*. 1st ed. Indianapolis: John Wiley and Sons.

International Telecommunication Union, Individuals Using the Internet: Australia (https://datahub.itu.int/data/?e=AUS&c=701&i=11624) (2024).

Kowalick, Phil, David Connery, and Rick Sarre. 2018. "Intelligence-Sharing in the Context of Policing Transnational Serious and Organized Crime: A Note on Policy and Practice in an Australian Setting." *Police Practice and Research* 19 (6): 596–608. https://doi.org/10.1080/15614263.2018.1507899.

National Anti-Scam Centre, National Anti-Scam Centre Advisory Board Terms of Reference (2023).

National Anti-Scam Centre, Scams Statistics Interactive Data (https://www.nasc.gov.au/scam-statistics) (2025).

RMIT University, Elon Musk Used in Fake AI Videos to Promote Financial Scam, available at https://www.rmit.edu.au/news/factlab-meta/elon-musk-used-in-fake-ai-videos-to-promote-financial-scam, last accessed on 28 March 2025.

Sheldon, Pavica. 2015. *Social Media: Principles and Applications*. Lanham : Boulder : New York : London: Lexington Books.

United Nations Office on Drugs and Crime, Casinos, Cyber Fraud, and Trafficking in Persons for Forced Criminality in Southeast Asia (2023).

9News, News, Sydney Bank Worker Loses $157,000 in 'Pig Butchering' Romance Scam, available at https://www.9news.com.au/national/sydney-bank-worker-loses-157000-in-pig-butchering-romance-scam/3fc559d6-163a-48c8-8317-283eb806a7c0 last accessed on 28 March 2025.

Legislations and Cases

Australian Anti-Money Laundering and Counter-Terrorism Financing Act 2006.

Australian Broadcasting Services Act 1992.

Australian Business Names Registration Act 2011

Australian Communications and Media Authority Act 2005.

Australian Competition and Consumer Act 2010.

Australian Corporations Act 2001.

Australian Financial Accountability Regime Act 2023.

Australian Financial Complaints Authority, Determination, Case number 12-00-1016692 (2024).

Australian Insurance Contracts Act 1984.

Australian Intelligence Services Act 2001.

Australian National Consumer Credit Protection Act 2009.

Australian Public Governance, Performance and Accountability Act 2013.

Australian Scam Prevention Framework Bill 2025 (as passed by both Houses version).

Australian Securities and Investments Commission Act 2001.

Australian Telecommunications Act 1997.

Federal Court of Australia, Judgment, Australian Competition and Consumer Commission v Qantas Airways Limited [2024] FCA 1219 (2024).

Online Fraud and Scams in India

Tuhinsubhra Giri

3

Online Fraud and Scams in India

Tuhinsubhra Giri[4]

BACKGROUND AND CONTEXT

The Rise of GenAI Online Scams and Implications

The digital revolution has ushered in an era where artificial intelligence (AI) is being weaponized by cybercriminals with alarming sophistication. Generative AI tools, which were originally designed to enhance creative and professional workflows, are now being exploited to orchestrate complex scams that are increasingly difficult to detect. According to Europol's 2025 SOCTA report, AI-driven scams are rapidly expanding in scale and sophistication, with deepfake technology and AI-generated phishing campaigns emerging as major threats to global cybersecurity. These scams often involve impersonating trusted entities, such as corporate executives, government officials, political leaders, celebrities or even family members, to manipulate victims into transferring money or divulging sensitive information. For instance, in 2023, a multinational company in Hong Kong lost $25 million after an employee was deceived by a deepfake video call featuring digitally recreated colleagues (South China Morning Post, 2024). Similarly, in the United States, AI voice cloning scams have surged, with criminals replicating the voices of loved ones to fabricate emergencies and extort money (Belanger, 2023). These cases underscore the global reach and adaptability of AI-powered fraud, highlighting the urgent need for robust countermeasures.

The proliferation of AI-driven scams is fueled by the accessibility of advanced tools. Open-source AI models and affordable cloud computing have democratized the ability to create convincing deepfakes and automated phishing schemes. Cybercriminals no longer require extensive technical expertise; instead, they can leverage user-friendly platforms to generate fraudulent content at scale. This trend is particularly concerning in regions with high digital penetration but limited cybersecurity awareness. For example, in Southeast Asia, AI-generated investment scams have proliferated on social media, luring victims with promises of unrealistic returns (Interpol, 2023). The global nature of these scams also complicates enforcement, as perpetrators often operate across jurisdictions, exploiting gaps in international cooperation. As AI technology continues to evolve, the threat landscape will likely expand, making it imperative for governments, businesses, and individuals to stay ahead of these emerging risks.

GenAI Online Scams in India

India, with its rapidly expanding digital economy, has become a prime target for AI-driven scams. The country's internet user base surpassed 800 million in 2023, making it the second-largest online market globally after China (India Foundation, 2025). This, combined with the explosive growth of digital payments, has created fertile ground for cybercriminals.

According to data presented in Parliament, Unified Payments Interface (UPI) fraud cases surged by 85% in FY24, rising from 7.25 lakh in FY23 to 13.42 lakh in FY24, with many involving phishing links and OTP theft (CNBC TV18, 2024). While the Reserve Bank of India (RBI) has not attributed a specific percentage

[4] Assistant Professor of Economics, Christ University

of these to AI, the growing use of AI-generated phishing and impersonation tactics has been widely acknowledged by cybersecurity experts.

One of the most alarming trends is the use of deepfake technology to spread misinformation and perpetrate fraud. During the 2024 Indian general elections, AI-generated videos of politicians making false or misleading statements circulated widely, raising concerns about their impact on democratic processes (The Hindu, 2024).

Another prevalent scam involves AI voice cloning, where fraudsters replicate the voices of family members to fake emergencies and extort money. In one case reported in March 2024, a Delhi-based businessman transferred ₹10 lakh after receiving a call from someone mimicking his son's voice using AI (Times of India, 2024).

India's cybersecurity infrastructure, while improving, struggles to keep pace with these advanced threats. The Indian Computer Emergency Response Team (CERT-In) has issued multiple advisories warning about deepfake scams and AI-generated fraud, including CIAD-2024-0060, which outlines threats and countermeasures (CERT-In, 2024). However, enforcement remains difficult due to the cross-border nature of many fraud operations. Scam call centers targeting Indian victims have been traced to countries like Cambodia and Myanmar, where jurisdictional limitations hinder effective crackdowns (Free Press Journal, 2023).

India's legal framework also lacks AI-specific provisions. The Information Technology (IT) Act, 2000, while foundational for cybersecurity, does not explicitly address generative AI misuse. The Digital Personal Data Protection (DPDP) Act, 2023 introduces important safeguards for data privacy, but it does not directly regulate AI-generated content or deepfake fraud (Chitranshi, 2023). This regulatory lag underscores the urgent need for policy to combat the escalating threat of AI-driven scams.

The economic ramifications of AI-driven scams are profound, particularly in a country like India, where digital financial inclusion is a key driver of growth. Indian Cyber Crime Coordination Centre (I4C), which reported ₹10,319 crore lost to online frauds between April 2021 and December 2023 (Times of India, 2024). However, the indirect costs are equally concerning. Scams erode public trust in digital platforms, discouraging adoption and stifling innovation. Small businesses and rural users, who are critical to India's digital transformation, may revert to cash transactions out of fear of fraud, undermining the government's efforts to promote a cashless economy. Furthermore, financial institutions and fintech companies are forced to invest heavily in AI-based fraud detection systems, driving up operational costs that are eventually passed on to consumers. These economic disruptions highlight the urgent need for systemic solutions to safeguard India's digital ecosystem.

AI-driven scams inflict significant social harm. Victims often experience psychological trauma, including anxiety, depression, and a lasting loss of trust in digital interactions. The societal impact is exacerbated by the targeting of vulnerable groups, such as the rural-urban digital divide, the elderly, and less tech-savvy individuals, who are disproportionately affected by voice cloning and phishing scams. Moreover, the spread of AI-generated misinformation, such as deepfake political content, threatens social cohesion and democratic processes. The erosion of trust in institutions, from banks to government agencies, poses a long-term challenge that extends beyond immediate financial losses.

India's Digital Landscape and Cybersecurity Vulnerabilities

India has experienced rapid digital transformation, driven by initiatives such as Digital India and the widespread adoption of the Unified Payments Interface (UPI). However, this digital expansion has also exposed the country to significant cybersecurity risks, including AI-driven fraud.

Some of the major Cybersecurity Vulnerabilities in India include;

- **Rise in AI-Powered Cyberattacks**

  A recent study indicates that 72% of Indian organizations have been targeted by AI-driven cyberattacks, including deepfake impersonation, phishing scams, and credential stuffing attacks (Free Press Journal, 2025; SME Futures, 2025).

- **Financial Fraud and Deepfake Scams**

  AI-enabled financial scams have reportedly caused losses exceeding ₹20,000 crore in 2024–25, with cybercriminals using deepfake videos to impersonate public figures and promote fake investment schemes (The Logical Indian, 2025).

- **Weak Cybersecurity Infrastructure**

  Despite progress in cybersecurity adoption, only 14% of Indian firms feel confident in defending against AI-based threats, while 36% say these threats surpass their existing detection tools (SME Futures, 2025).

- **Lack of AI-Specific Regulations**

  India's legal framework, including the Information Technology Act of 2000, does not fully address the challenges posed by AI-generated fraud. The country lacks comprehensive laws focused on generative AI misuse and cross-border enforcement (LawArticle, 2025).

Technology Adoption Lifecycle & The Evolution of AI-Driven Scams

AI-driven scams evolve in tandem with the Technology Adoption Lifecycle.

- This process begins in fringe cybercrime circles before scaling into mainstream fraud tactics.
- As tools become easier to use, the early majority phase sees mass deployment of phishing bots and synthetic identity fraud.
- In the late majority phase, scams become widespread, prompting institutional countermeasures.

This progression mirrors the disruptive innovation model, where new AI-enabled fraud replaces conventional tactics. A prime example is how AI-generated investment scams have overtaken manual Ponzi schemes, thanks to their scalability and realism.

The fraud tactics have evolved significantly over the past decade (see annex A). Before 2018, scams relied on basic impersonation calls, generic phishing emails, and Ponzi schemes. However, from 2019 onward, AI has revolutionized financial fraud—deepfake videos now impersonate trusted figures, AI-generated phishing messages exploit personal data, and synthetic identities automate loan fraud. Scams have become hyper-realistic, scalable, and harder to detect, with AI enabling real-time deception through chatbots, cloned customer support, and fake e-commerce sites. The shift from human-led cons to AI-driven fraud underscores the urgent need for advanced detection tools and regulatory frameworks to combat these sophisticated threats. How Generative AI can be used or exploited to facilitate or enhance each scam can be seen in Annex B.

PATTERNS AND TRENDS OF GEN AI-DRIVEN ONLINE SCAMS IN INDIA

The Digital Surge and Its Dark Underside

India's rapid digital transformation, fueled by Digital India, UPI, and Aadhaar-based services, has inadvertently created new vulnerabilities for Generative AI-driven scams. Fraudsters now exploit AI to generate synthetic identities, bypassing traditional verification systems. AI-generated fake Aadhaar and PAN cards are being used to access financial services fraudulently, leading to unauthorized loans and identity theft (Economic Times, 2024). Additionally, deepfake technology enables real-time impersonation, allowing scammers to mimic bank officials or government representatives in video calls to deceive victims into transferring money. The Aadhaar-enabled Payment System (AePS) has seen a rise in fraud cases, with Aadhaar breaches in land records contributing to the surge in AePS fraud, exploiting fingerprint cloning and duplicate biometric records (Medianama, 2024).

The widespread adoption of UPI has also made AI-powered scams more sophisticated. Reports indicate that 55% of digital payment frauds in India are linked to UPI, with AI-driven phishing attacks targeting unsuspecting users (Business Standard, 2023). Fraudsters use AI-generated voice cloning to impersonate family members in distress, coercing victims into sending money. AI-powered chatbots further automate scam operations, responding dynamically to victims' queries and making fraudulent schemes appear more legitimate. As AI tools become more accessible, scams are evolving beyond simple phishing attempts into highly personalized fraud campaigns, making detection increasingly difficult. Without stronger AI-specific regulations, India's digital economy remains vulnerable to large-scale financial fraud.

Recent Statistics of Financial Losses and Reported Cases (2020-2025)

India has witnessed a dramatic surge in cybercrime over the past five years, driven by increased digital adoption and the proliferation of AI tools. According to the Future Crime Research Foundation (FCRF), online financial fraud accounted for 77.41% of all cybercrime cases reported between January 2020 and June 2023, making it the most dominant category of cybercrime in India[5].

Another major problem with this is that the rapid digitalisation of the economy also attracts this kind of scam and problems. Cybercrime cases themselves are not very old, Gen AI-related online scams are quite new in comparison. But there is a problem with the availability of data for this problem. Some data on cybercrime is available, but not segregated as Gen AI related financial scams. Though this is one type of Cybercrime, so data on cybercrime can definitely give us some idea about the increasing number of Gen AI-related financial scams.

Table 3.1 Number of Cybercrime Cases and Estimated Financial Loss Over the Period 2020-2024

Year	Reported Cybercrime Cases	Estimated Financial Loss (₹ crore)	Source
2020	50,035	1,785	Economic Times, 2024
2021	52,974	2,096	Statista, 2024
2022	65,893	3,192	Business Standard, 2024

[5] https://the420.in/fcrf-cybercrime-report-india-77-percent-online-financial-fraud/

2023	76,630	4,820	Times of India, 2024
2024 (est.)	89,000+	6,500+	Medianama, 2024

Between January and April 2024, India recorded over 740,000 cybercrime complaints, as reported by the Indian Cyber Crime Coordination Centre (I4C). This marks a significant increase compared to previous years, with cybercrime cases surging between 2019 and 2020 and continuing to rise steadily. Notably, approximately 85% of these complaints in 2024 were linked to online financial fraud, including investment scams, illegal lending apps, and phishing attacks (Statista, 2024).

Cross-Border Scam Activities and India's Exposure

India is increasingly vulnerable to transnational AI-driven scams, especially those involving cryptocurrency, investment fraud, and identity theft. These scams often originate from jurisdictions with weak enforcement and exploit global platforms like WhatsApp, Telegram, and fake websites.

- Cross-border UPI frauds and SIM swap attacks have been traced to Southeast Asia and Eastern Europe.
- AI-generated deepfakes are used to bypass KYC protocols, enabling money laundering and mule account creation
- According to CloudSEK, brand impersonation and cross-border phishing campaigns are among the top threats to Indian financial institutions.

India's regulatory and enforcement agencies, including CERT-In, RBI, and NCRB, are increasingly collaborating with global cybersecurity firms to track and mitigate these threats. However, the lack of harmonized international frameworks continues to hinder effective prosecution and recovery. India is a prime target for transnational AI fraud networks, with 70% of scam calls originating from Cambodia, Myanmar, and Laos[6].

Table 3.2 Transnational Scam Operations Targeting Indian Victims in Recent Times

Country	Scam Particulars	Primary Scam Types	Indian Victims	Source
Cambodia	100,000 scammers generating an estimated $12.8 billion in 2013	AI Voice Cloning, Fake Job Scams	More than 5,000	Voanews, Indiana-express
Myanmar	About 120,000 individuals being forced into scamming in Myanmar in the last year	Invest-ment Frauds, Romance Scams	549 Indians were freed from cybercrime centres in Myanmar-Thailand border	BBC, Hindu-stantimes
Laos	306 call centers or fraud units are identified in SEZs	UPI Fraud, Fake Customer Support	To date, 924 Indian nationals have been rescued	Laotian-times

[6] https://laotiantimes.com/2024/10/29/india-faces-rising-digital-scams-linked-to-laos-myanmar-cambodia/

China	Transnatio-nal Criminal networks from China dominate Southeast Asia's gambling and scam operations	Deepfake Video Scams, Loan Frauds	The NIA reports that many Indians were recruited via fraudulent job ads and compelled to work under coercive contracts	Usip, Indiato-day

Source: Author's compilation

Southeast Asia has become a major hub for transnational scam operations. As we can see from Table above that, in Cambodia alone, 100,000 scammers generated an estimated $12.8 billion in fraudulent activity, nearly half the country's GDP. Myanmar's scam networks, often linked to criminal syndicates, force over 120,000 individuals into scams such as crypto fraud and romance-investment scams, with 549 Indians rescued from cybercrime centres near the Thai border. Laos' Golden Triangle Special Economic Zone hosts 306 scam units, where Indian nationals are trafficked into forced cyber fraud-924 Indians have been rescued to date. Many times, China's networks dominate global gambling and online fraud, with $64 billion stolen annually, as syndicates exploit a $40-$80 billion market. Reports indicate many more scams remain unaccounted for, highlighting the urgent need for cross-border regulatory enforcement and digital fraud prevention strategies.

Case Studies

CBI's Operation Chakra-V (2025)

Operation Chakra-V, launched by the Central Bureau of Investigation (CBI) in 2025, marked a significant milestone in India's fight against cyber-enabled financial fraud. The operation targeted an international cybercrime syndicate that exploited AI-driven scams, including spoofed caller IDs, deepfake impersonation, and voice cloning to deceive victims, primarily in the United States and Canada. Investigators uncovered fraudulent operations linked to tech-support scams, impersonation of government officials, and cryptocurrency laundering schemes. During coordinated raids across three locations in India, CBI seized ₹2.8 crore in cryptocurrency, ₹22 lakh in cash, and multiple fake digital identities, disrupting a complex fraud network.

Beyond financial fraud, Operation Chakra-V highlighted the growing role of AI in cybercrime and the urgent need for AI-specific cybersecurity regulations. The operation also strengthened international cooperation, with CBI working alongside Interpol and the FBI to trace global money trails. Following the arrests, India's cyber enforcement agencies intensified efforts to strengthen digital forensic capabilities and fraud detection frameworks. With AI-powered scams evolving rapidly, Operation Chakra-V underscores the importance of proactive cyber-defense strategies in safeguarding India's digital economy.

AI-Generated Flipkart Scam (2023)

The AI-Generated Flipkart Scam in 2023 exemplifies the growing sophistication of AI-driven cyber fraud in India's digital economy. Fraudsters leveraged generative AI to clone Flipkart's website, creating a near-identical replica that deceived thousands of unsuspecting shoppers. By running fake discount campaigns, scammers lured over 30,000 victims into purchasing non-existent products, leading to an estimated ₹120 crore in financial losses. The scam exploited AI-generated phishing tactics, including automated customer service bots and deepfake promotional videos, making detection difficult until significant damage had been done.

This case highlights the urgent need for AI-specific cybersecurity regulations and enhanced fraud detection mechanisms. As AI tools become more accessible, cybercriminals are increasingly using synthetic identities, deepfake impersonation, and automated scam operations to bypass traditional security measures. The Flipkart scam underscores the importance of digital literacy, real-time fraud monitoring, and stricter e-commerce verification protocols to protect consumers from AI-enabled deception.

The Cambodia Cyber Scam Factories (2023)

In 2023, a major cyber scam operation in Cambodia exposed the forced involvement of Indian nationals in fraudulent online schemes. According to reports, over 5,000 Indians were coerced into working in scam centers, where they were made to conduct online fraud, including money laundering, crypto scams, and romance fraud. Victims were initially lured with fake job offers, only to find themselves trapped in illegal cyber operations upon arrival.

The Indian government intervened, successfully rescuing 250 citizens and working closely with Cambodian authorities to crack down on the scam networks. The case highlights the growing transnational nature of cyber fraud, where AI-driven deception tactics—such as deepfake impersonation and automated phishing—are increasingly used to exploit victims. It also underscores the urgent need for international cooperation and AI-specific cybersecurity regulations to prevent such large-scale fraud operations.

IMPLICATIONS OF ONLINE FRAUD AND SCAMS

Financial Impact on Individuals, SMEs, and the Economy

AI-driven online scams have inflicted significant financial damage across all segments of Indian society. According to a 2025 report by The Logical Indian[7], India is projected to lose over ₹20,000 crore to AI-enabled scams in a single year, with deepfake investment frauds and impersonation scams being the primary culprits. These scams often impersonate public figures like Finance Minister Nirmala Sitharaman or Google CEO Sundar Pichai to promote fake investment platforms, leading to widespread deception and monetary loss.

For individuals, especially those with limited digital literacy, the financial consequences can be devastating. Victims often lose life savings or emergency funds, with little recourse for recovery. Small and medium enterprises (SMEs) are also vulnerable, particularly to phishing attacks and synthetic identity fraud. A study by Experian and Forrester Consulting found that 64% of Indian financial institutions reported increased fraud losses in 2024, with synthetic identity fraud being the most prevalent[8].

At the macroeconomic level, the cumulative effect of these scams undermines investor confidence, increases cybersecurity costs, and diverts resources from productive sectors. These kind of Gen AI Online scams leads to business disruptions, intellectual property theft, and increased expenditure on fraud prevention, all of which hamper economic growth.

[7] https://thelogicalindian.com/india-faces-%E2%82%B920000-crore-cybercrime-threat-in-2025-amid-surge-in-ai-driven-deepfake-investment-scams/

[8] https://www.experian.in/2024/02/11/financial-frauds-rise-in-india-as-genai-gains-traction/

Psychological and Social Impact

The psychological toll of AI-driven scams is often overlooked but deeply consequential. Victims experience a range of emotional responses, including shock, anxiety, shame, and depression. A 2023 article in The Times of India[9] highlights how deepfake scams and voice cloning can cause lasting mental distress, especially when victims are manipulated into believing a loved one is in danger.

The emotional manipulation involved in AI scams—such as receiving a cloned voice call from a family member in distress—can lead to trauma and long-term distrust. Victims often suffer from self-blame, fear of future scams, and social withdrawal, particularly when the scam involves sextortion or impersonation. These effects are compounded in cases involving adolescents or the elderly, who may lack the tools or support systems to recover emotionally.

Digital Trust Erosion in Financial Institutions and E-Commerce Platforms

AI scams have significantly eroded public trust in digital platforms. These scams not only cause financial loss but also undermine the credibility of financial institutions and e-commerce platforms.

A survey by Finextra found that 63% of Indian consumers have either fallen victim to a scam or know someone who has, leading many to reduce their use of digital payment platforms[10]. This erosion of trust has broader implications: it slows down digital adoption, increases reliance on cash transactions, and hampers the growth of India's fintech ecosystem.

Consumer behaviour is also shifting. According to FICO's 2025 India Fraud Report, users are becoming more cautious, often avoiding online transactions or demanding additional verification steps. While this may enhance security, it also reduces the convenience and efficiency that digital platforms are designed to offer.

Government and Institutional Regulatory Responses

The surge in AI scams has forced policymakers to accelerate regulatory reforms. Some of the key actions include:

- **Digital Personal Data Protection Act (DPDP), 2023** – Introduces penalties for misuse of personal data in AI fraud but lacks specific provisions on deepfakes.
- **RBI's AI Fraud Prevention Guidelines (2024) –** Mandates banks to deploy AI-based deepfake detection and multi-factor authentication for high-risk transactions.
- **Interpol-India Collaboration –** Targeting offshore scam hubs in Cambodia and Myanmar, leading to 50+ arrests in 2024.

[9] https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/the-deep-impacts-of-deepfakes-and-cyber-fraud-on-mental-health/articleshow/106145692.cms

[10] https://www.finextra.com/blogposting/27108/digital-arrest-a-new-frontier-in-cybercrime-and-its-ripple-effects-on-consumer-trust

Table 3.3 Sectoral Impact-Who Is Being Targeted?

Sector	Primary AI-Driven Threats	Impact
Banking & Fintech	Deepfake investment scams, synthetic identity fraud	Loss of consumer trust, regulatory fines
Retail & E-Commerce	Fake websites, chatbot impersonation	Brand damage, customer losses
Government Services	Deepfake impersonation of officials	Public misinformation, reputational harm
Telecom	Phishing via SMS/WhatsApp	SIM swap fraud, identity theft

Other than the individual level impact, AI-driven scams have significantly impacted critical industries, eroding trust, financial stability, and operational integrity. From the table 4 above, we can see that the banking and fintech sector faces rising fraud cases through deepfake investment scams and synthetic identity fraud, forcing tighter regulatory oversight. Retail and e-commerce platforms suffer brand damage and customer losses due to fake websites and chatbot impersonation, making digital trust harder to maintain. In government services, deepfake impersonation of officials fuels public misinformation, threatening policy credibility. Meanwhile, the telecom sector experiences phishing attacks via SMS and WhatsApp, leading to SIM swap fraud and identity theft, amplifying security concerns across digital transactions. These sectoral vulnerabilities demand AI-specific cybercrime laws, rules, regulations, and robust AI-driven fraud detection strategies and heightened consumer awareness initiatives.

KEY STAKEHOLDERS IN ADDRESSING AI-DRIVEN SCAMS IN INDIA

The rise in AI-driven scams has prompted not only government, but also other stakeholders to launch initiatives aimed at addressing scams. Below is an overview of the key actors and their efforts to address online fraud and scams in India.

Government Bodies

Government agencies play a pivotal role in regulating, preventing, and mitigating AI-driven scams.

Stakeholder	Key Responsibilities	Recent Actions (2023-24)
CERT-In (Indian Computer Emergency Response Team)	• National nodal agency for cybersecurity threats • Issues alerts on AI scams • Coordinates with ISPs to block fraudulent domains	• Launched AI-powered scam tracking system (2024) • Reported 12,000+ deepfake fraud cases in 2023
Ministry of Electronics & IT (MeitY)	• Formulates AI and cybersecurity policies • Regulates digital platforms	• Released AI Ethics Guidelines (2024) • Proposed ban on malicious deepfakes
Reserve Bank of India (RBI)	• Safeguards financial systems from AI fraud • Mandates fraud detection for banks	• Introduced AI-based UPI fraud detection (2024 • Reported ₹23,000 crore in AI banking scams (2023)
Securities and Exchange Board (SEBI)	• Prevents stock market fraud via AI • Monitors fake investment schemes	• Banned 350 AI-powered trading scams (2024)

Private Sector Participation

Private companies, especially in fintech and telecom, are crucial in detecting and preventing scams.

Sector	Key Players	Anti-Scam Measures
Fintech & Digital Payments	Paytm, PhonePe, NPCI	• AI-based transaction anomaly detection • Real-time fraud alerts via SMS/email
Telecom Providers	Jio, Airtel, Vodafone-Idea	• AI call monitoring to flag scam numbers • Blocked 10M+ spam calls monthly (TRAI, 2024)
E-Commerce & Social Media	Flipkart, Amazon, Meta	• Deepfake detection algorithms • Verified seller programs

Tech & AI Companies

Global tech giants and Indian startups are deploying AI to counter AI-driven fraud.

Company	Role in Scam Prevention	Key Initiatives
Google	• Detects phishing sites • Flags scam ads	• Deepfake watermarking in Google Search
Microsoft	• Azure AI for fraud detection • Secure digital identities	• AI voice clone detection for banks
Indian Startups (e.g., SigTuple, RazorpayX)	• AI-based KYC fraud prevention • Scam pattern recognition	• Reduced fraud by 40% in partner banks (2024)

Law Enforcement & Cybersecurity Firms

Cybercrime units and cybersecurity firms track and dismantle scam operations.

Agency/Firm	Function	Notable Cases (2023-24)
Indian Cyber Crime Coordination Centre (I4C)	• Tracks transnational scam networks • Trains police in AI fraud detection	• Busted Cambodia-based AI call center scamming Indians
Delhi/Mumbai Cyber Cells	• Investigates financial fraud • Recovers stolen funds	• Solved ₹5.7 crore AI voice scam (2023)
Cybersecurity Firms (e.g., Kaspersky, Quick Heal)	• Develop AI scam detection tools • Provide threat intelligence	• Blocked 5M+ phishing attempts in India (2024)

POLICY ASSESSMENT

Existing Laws & Regulations

India has several legal frameworks addressing cybersecurity and financial fraud, but they lack AI-specific provisions. The following laws and regulations play a role in mitigating online scams.

Information Technology (IT) Act, 2000 and Amendments

The IT Act, 2000 is India's primary legislation governing cybercrime and electronic commerce. It provides legal recognition for digital transactions and penalizes cyber fraud. However, the Act does not explicitly address AI-driven scams, deepfake fraud, or synthetic identity manipulation. Amendments have been made to strengthen cybersecurity, but AI-driven fraud detection remain absent.

RBI Guidelines on Financial Fraud

The Reserve Bank of India (RBI) has issued multiple guidelines to combat financial fraud, including:

- Master Directions on IT Governance (2023), which mandate banks and financial institutions to adopt AI-driven fraud detection systems.
- Cybersecurity Framework for Banks, requiring real-time monitoring of digital transactions.
- Guidelines on Digital Lending, aimed at preventing fraudulent loan applications using synthetic identities.

Despite these measures, AI-powered scams continue to exploit loopholes in digital banking security.

Consumer Protection Laws Relevant to Digital Transactions

The Consumer Protection Act, 2019 and the Digital Personal Data Protection Act, 2023 provide safeguards against fraudulent digital transactions. The Central Consumer Protection Authority (CCPA) oversees deceptive practices, but enforcement against AI-generated scams remains weak. AI-generated misinformation and fraudulent e-commerce platforms often bypass existing consumer protection mechanisms.

Challenges

Despite existing regulations, several challenges hinder effective enforcement against AI-driven scams.

a) Broader definitions of scams to cover AI-driven

India lacks dedicated AI-driven scams provisions in the existing law to regulate deepfake fraud, AI-generated phishing, and synthetic identity scams. While the IndiaAI Mission (2024) aims to develop ethical AI frameworks, it does not directly address AI-driven cybercrime. The absence of AI-specific liability frameworks makes it difficult to prosecute fraudsters using AI tools. There is a need to incorporate AI-driven scams in the under the current IT Act.

b) Limited Cross-Border Enforcement Mechanisms

AI-driven scams often originate from international cybercrime syndicates, making enforcement difficult. India's cybercrime laws do not have strong cross-border provisions, limiting cooperation with global agencies. The lack of extradition treaties for cybercriminals further complicates prosecution. A huge number of AI scam operations are run from Cambodia, Myanmar, and Laos, as mentioned earlier. Mutual Legal Assistance Treaties (MLATs) with these countries are often slow, allowing scam networks to evade shutdowns.

c) Inadequate Digital Literacy Among Users

A major challenge in combating AI scams is low digital literacy. Only 38 percent of households in the country are digitally literate. Additionally, only 31 percent of the rural population uses the

internet as compared to 67 percent of the urban population[11]. Many users are unaware of AI-driven fraud tactics, making them vulnerable to scams. Public awareness campaigns on AI fraud detection are needed to bridge this gap.

India's policy landscape for addressing AI-driven scams is evolving but remains fragmented and reactive. Strengthening IT Act by including AI-driven scams provisions, enhancing cross-border enforcement, and improving digital literacy are crucial steps toward mitigating AI-enabled fraud. The government must integrate AI governance frameworks into cybersecurity laws to ensure a proactive approach to fraud prevention.

Table 3.4. Comparative Policy Analysis of India and its Asia-Pacific Peers

Policy Aspect	India	China (Interim AI Measures, 2023)	Japan (AI Governance Guidelines, 2024)	Singapore (AI Verify Framework)	Australia (AI Ethics Principles)
AI Scam Definition	No explicit classification	Strict AI content labelling, bans unauthorized deepfakes	Ethical AI guidelines, voluntary compliance	AI risk-based classification	Ethical AI guidelines, no legal mandate
Cross-Border Cooperation	Limited MLAT effectiveness	Cybersecurity cooperation with ASEAN	International AI safety partnerships	ASEAN cybersecurity cooperation	International AI safety partner-ships
Public Awareness	Ad-hoc campaigns (e.g., Cyber Jaagrookta Diwas)	AI literacy in schools, public misinformation monitoring	AI ethics education on in universities	AI literacy initiatives	AI ethics education in universities

BEST POLICY PRACTICES FOR GENAI ONLINE SCAM PREVENTION

AI Transparency and Content Labelling

European Union – AI Act & Digital Services Act (DSA)

The EU mandates labelling of AI-generated content, bans high-risk applications like unauthorized deepfakes, and requires platforms to verify advertisers. Non-compliance can result in fines up to 6% of global revenue.

China – Interim Measures for Generative AI (2023)

Requires synthetic content to be labelled, prohibits impersonation without consent, and mandates that AI-generated content must not endanger national security or social stability.

Cross-Border Intelligence Sharing

Interpol & Europol Joint Task Forces

Facilitate real-time data exchange on transnational scams, including AI-enabled fraud. These platforms support coordinated takedowns and intelligence-led investigations.

[11] https://idronline.org/article/inequality/indias-digital-divide-from-bad-to-worse/

ASEAN Cybersecurity Cooperation

Southeast Asian nations collaborate on cross-border scam prevention, sharing threat intelligence and harmonizing digital fraud response protocols.

Public Awareness and Digital Literacy

Japan – AI Ethics in Education

Integrates AI literacy and scam awareness into school curricula and public campaigns to build early digital resilience.

Australia – eSafety Commissioner Initiatives

Runs national campaigns on deepfake awareness, scam reporting, and content takedown protocols, modelled on the Online Safety Act.

Private Sector and Regulatory Collaboration

United States – FTC & NIST AI Risk Management Framework

The U.S. Federal Trade Commission (FTC) uses AI to analyze consumer complaints and detect scam patterns. The NIST AI Risk Management Framework promotes red-teaming and stress-testing of AI systems to reduce vulnerabilities.

Google's Global Scam Policy Recommendations

Advocates for cross-sector collaboration, real-time scam intelligence sharing, and proactive takedown of malicious content. Google's pilot in Singapore blocked nearly 900,000 high-risk app installations.

AI-Specific Legal Frameworks

Singapore – Model AI Governance Framework & Anti-Scam Command (ASCom)

Combines real-time scam detection, mandatory SMS sender ID registration, and public-private coordination. Phishing losses dropped by 37% in 2023.

UK – Online Fraud Charter (2023)

Requires banks, telecoms, and tech firms to share fraud data within 24 hours, improving scam response time and consumer protection.

POLICY RECOMMENDATIONS

- **AI Scam Registry & Shared Intelligence Grid**

  Create a centralized fraud intelligence platform that banks, fintech firms, telecoms, e-commerce companies, and law enforcement can access. Patterned after the UK's National Fraud Database, it would enable real-time sharing of scam indicators like deepfake voiceprints, scam URLs, and synthetic identity hashes.

- **AI Media Provenance & Content Labeling Mandate**

  Introduce legislation requiring mandatory watermarking and cryptographic labeling of AI-generated images, videos, and voices. This will bolster public trust and help platforms automatically flag deceptive content before dissemination.

- **Sector-Specific AI Risk Certification for Platforms**

  Inspired by the U.S. NIST Framework, mandate that high-risk sectors (e.g., banking, telecom, digital advertising) undergo annual AI fraud risk audits, including red-teaming, explainability testing, and consumer impact simulations. CERT-In and RBI could co-lead certification.

- **AI Scam Literacy in Government Portals & Education**

  Integrate regional-language chatbot explainers about AI scams into key portals like MyGov, DigiLocker, and PMGDISHA. Simultaneously, embed scam resilience modules into CBSE/NCERT digital literacy curriculum.

- **National Rapid Takedown Protocol**

  Enact a cross-sector protocol requiring platforms to takedown verified scam content (e.g., deepfake investment ads) within 6 to12 hours of verified flagging by I4C, CERT-In, or RBI. This protocol can emulate Australia's eSafety takedown standard.

- **National Scam Simulation Challenge**

  Launch an annual innovation challenge for universities, startups, and police academies to prototype:
  - Deepfake detection models
  - Real-time scam alert apps
  - Local-language scam literacy games

  Winners could receive funding and regulatory fast-tracking for national deployment, catalyzing homegrown, scalable scam-tech solutions.

- **Establish a Cross-Border Generative AI Scam Intelligence Taskforce (GEN-SAFE*)***

  India should initiate or co-lead a multilateral taskforce called GEN-SAFE (Generative AI Scam and Fraud Exchange) in collaboration with ASEAN, Interpol, and select G20 digital economy members.

- **Public-Private Collaboration Frameworks**

  A National AI Scam Registry should be created as a unified database of scam signatures, integrating with banking systems, e-commerce platforms, and telecom providers.

- **Consumer Empowerment Initiatives**

  The government can launch a ScamScan app with UPI verification, deepfake detection, and one-click police reporting. A digital literacy offensive must make "AI Spotting" modules mandatory in schools and Digital India centers, while expanding the MyGov "FRAUD AI" chatbot that already served 2.1 million users in 12 languages during its pilot phase.

ANNEXES

Annex A. Evolution of Financial Scams in India – From Traditional to AI-Driven

Category	Traditional Financial Scams (Pre-2018)	AI-Driven Online Scams (2019–2025)	Key Changes
Impersonation Scams	Fake calls from bank officials or police demanding KYC updates or legal payments.	Deepfake videos and voice cloning of public figures (e.g., Finance Minister, Sundar Pichai) promoting fake investment platforms.	Shift from human-led deception to AI-generated hyper-realistic impersonation.
Phishing Attacks	Generic emails or SMS with suspicious links.	AI-generated personalized phishing emails and WhatsApp messages using NLP and behavioural data.	Enhanced targeting and believability due to AI's ability to mimic tone and context.
Investment Scams	Ponzi schemes, chit funds, or fake stock tips.	AI-generated fake websites and social media ads with deepfake endorsements for crypto or government bonds.	Use of generative AI to simulate legitimacy and scale outreach.
Loan & Credit Scams	Fake loan offers via SMS or calls.	AI-created synthetic identities used to apply for loans or open mule accounts.	Automation of identity fraud using AI-generated documents and profiles.
Customer Support Fraud	Fake helpline numbers or spoofed IVR systems.	AI chatbots mimicking bank or e-commerce support to extract OTPs and credentials.	Real-time AI interaction increases success rate of deception.
E-Commerce Fraud	Non-delivery of goods from fake websites.	AI-generated scam websites mimicking trusted brands with cloned UI and fake reviews.	AI enables rapid creation of convincing fraudulent storefronts.
Romance & Sextortion Scams	Fake dating profiles using stolen photos.	AI chatbots and deepfake avatars used to build emotional trust and extort money.	Emotional manipulation scaled through generative AI and voice cloning.
Cross-Border Fraud	Email lottery scams or Nigerian prince frauds.	Transnational AI scams using multilingual phishing, crypto laundering, and global mule networks.	AI enables cross-border scalability and evasion of local enforcement.

Annex B. Misuse of AI for Online Scams in India

Type of Scams	Cases in India	Source
Phishing scams	Indians receive an average of 12 fake messages per day, many impersonating banks or government agencies	The Hindu – AI-powered phishing surge in India
Investment fraud	Deepfake videos of Finance Minister Nirmala Sitharaman and Google CEO Sundar Pichai were	NDTV – False Endorsements, Real Losses

	used to promote fake crypto platforms like "InvestGPT."	
Fake job offers	A Chennai woman was targeted by an AI chatbot posing as a recruiter from "Flypside Global Services," offering a fake job via WhatsApp	PyLessons – Bot-Driven Job Scams
Loan app scams	RBI flagged over 1,200 fake loan apps in Q1 2025; victims were harassed and extorted after borrowing. Generative AI created fake content and deepfaked video 'loan officers' also being used for this kind of scams.	OneTouch Finance – Fake Loan App List
Deepfake scams	A 73-year-old man in Kerala lost ₹40,000 after a scammer used a deepfake video call to impersonate his former colleague. The scammer used AI-enabled deepfake technology to create a video call in which the impersonator's face and voice matched the victim's former colleague. This is the first reported case of a deepfake scam	Hindustan Times – Kerala Deepfake Scam
Online shopping fraud	In 2024, Indians lost over ₹11,000 crore to online fraud, including fake e-commerce sites using AI-generated product images and reviews	ET CIO – AI in Fintech Fraud Prevention
OTP and UPI scams	A Mumbai resident lost ₹95,000 in 3 minutes after sharing an OTP with a scammer posing as a bank official	Sprouts News – UPI Verification Scam
Romance scams	A Bengaluru woman lost ₹5.4 lakh to a scammer posing as an army officer on Tinder, who used emotional manipulation and AI-generated content	Media India – AI & Romance Scams
Fake digital arrest threats	A Mumbai professional was coerced into paying ₹50,000 after a scammer impersonated an income tax officer and threatened "digital arrest."	Union Bank of India – Digital Arrest Case
TRAI (Telecom Regulatory Authority of India) Impersonation Scam	An elderly woman in Chandigarh lost ₹2.5 crore after scammers impersonated TRAI and CBI officials, threatening to disconnect her phone number	Indian Express – TRAI Impersonation Scam

REFERENCES

Belanger, A. (2023, March 6). *Thousands scammed by AI voices mimicking loved ones in emergencies*. Ars Technica. https://arstechnica.com/tech-policy/2023/03/rising-scams-use-ai-to-mimic-voices-of-loved-ones-in-financial-distress/

Europol. (2025). *EU Serious and Organised Crime Threat Assessment (EU-SOCTA) 2025*. European Union Agency for Law Enforcement Cooperation. Retrieved from https://www.europol.europa.eu/publications-events/main-reports/socta-report

INTERPOL. (2023, December 8). *INTERPOL operation reveals further insights into 'globalization' of cyber scam centres*. Retrieved from https://www.interpol.int/en/News-and-Events/News/2023/INTERPOL-operation-reveals-further-insights-into-globalization-of-cyber-scam-centres

South China Morning Post. (2024, February 4). *'Everyone looked real': Multinational firm's Hong Kong office loses HK$200 million after scammers stage deepfake video meeting*. Retrieved from https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage

CERT-In. (2024, March 8). *Advisory on threats posed by deepfakes powered by artificial intelligence and related countermeasures (CIAD-2024-0060)*. Indian Computer Emergency Response Team. https://www.cert-in.org.in/s2cMainServlet?pageid=PUBVLNOTES02&VLCODE=CIAD-2024-0060

CNBC TV18. (2024, June 5). *UPI fraud cases rise 85% in FY24 to 13.4 lakh: Parliament reply*. https://www.cnbctv18.com/business/finance/upi-fraud-cases-rise-85-pc-in-fy24-increase-parliament-reply-data-19514295.htm

Free Press Journal. (2023, October 24). *CERT-In issues advisory on AI-powered deepfakes, warns citizens of scammers using realistic tactics for financial fraud*. https://www.freepressjournal.in/mumbai/cert-in-issues-advisory-on-ai-powered-deepfakes-warns-citizens-of-scammers-using-realistic-tactics-for-financial-fraud

India Foundation. (2025, April 15). *Fortifying the digital frontier: Protecting India's cyber interests*. https://indiafoundation.in/articles-and-commentaries/fortifying-the-digital-frontier-protecting-indias-cyber-interests/

Chitranshi, S. (2023, September 7). The deepfake conundrum: Can the Digital Personal Data Protection Act, 2023 deal with misuse of generative AI? *Indian Journal of Law and Technology* (IJLT). https://www.ijlt.in/post/the-deepfake-conundrum-can-the-digital-personal-data-protection-act-2023-deal-with-misuse-of-ge

The Hindu. (2024, April 16). *From IT bots to AI deepfakes: The evolution of election-related misinformation in India*. https://www.thehindu.com/elections/lok-sabha/from-it-bots-to-ai-deepfakes-the-evolution-of-election-related-misinformation-in-india/article68015342.ece

Times of India. (2024, March 17). *Fooled by your own kid? Chilling rise of AI voice cloning scams*. https://timesofindia.indiatimes.com/india/fooled-by-your-own-kid-chilling-rise-of-ai-voice-cloning-scams/articleshow/108569446.cms

Times of India. (2024, January 4). *India saw 129 cybercrimes per lakh population in 2023*. Retrieved from https://timesofindia.indiatimes.com/india/india-saw-129-cybercrimes-per-lakh-population-in-2023/articleshow/106524847.cms

Free Press Journal. (2025, June 9). *Cybercrime alert: 72% Indian organisations targeted; AI becomes new weapon enabling stealthier attacks*. https://www.freepressjournal.in/business/cybercrime-alert-72-indian-organisations-targeted-ai-becomes-new-weapon-enabling-stealthier-attacks

India Foundation. (2025, May 1). *Fortifying the digital frontier: Protecting India's cyber interests*. https://indiafoundation.in/articles-and-commentaries/fortifying-the-digital-frontier-protecting-indias-cyber-interests

LawArticle. (2025, June 11). *Emerging cybercrime and the AI impact*. https://lawarticle.in/emerging-cybercrime-and-the-ai-impact

SME Futures. (2025, June 9). *72% Indian firms hit by AI-powered cyberattacks in past year: Report*. https://smefutures.com/72-indian-firms-hit-by-ai-powered-cyberattacks-in-past-year-report

The Logical Indian. (2025, June 5). *India faces ₹20,000 crore cybercrime threat in 2025 amid surge in AI-driven deepfake investment scams*. https://thelogicalindian.com/india-faces-%E2%82%B920000-crore-cybercrime-threat-in-2025-amid-surge-in-ai-driven-deepfake-investment-scams

Economic Times. (2024, May 10). *How AI-generated Aadhaar and PAN card frauds are rising*. Retrieved from https://economictimes.indiatimes.com

Business Standard. (2023, May 16). *UPI-related scams account for 55% of total digital payments frauds in India*. Retrieved from https://www.business-standard.com/finance/news/upi-related-scams-account-for-55-of-total-digital-payments-frauds-in-india-123051600333_1.html

Medianama. (2024, July 7). *Aadhaar breaches in land records behind AePS fraud surge*. Retrieved from https://www.medianama.com/2024/07/223-aadhaar-breaches-in-land-records-behind-aeps-fraud-surge/#:~:text=AePS'%20contribution%20to%20financial%20fraud,fingerprint%20and%20a%20duplicate%20one

Statista. (2024). *Cybercrime cases reported to I4C India (2019–2024)*. Retrieved from https://www.statista.com/statistics/1499739/india-cyber-crime-cases-reported-to-i4c/

Online Fraud and Scams in Indonesia

Adinova Fauri, Futy Ichiradinda,
Rojwa Rachmiadi

4

Online Fraud and Scams in Indonesia

Adinova Fauri[12], Futy Ichiradinda[13], Rojwa Rachmiadi[14]

INTRODUCTION

Over the past decade, the internet penetration rate in Indonesia has more than doubled from 88,1 million users in 2014 to nearly 221.6 million users or 79,5 percent of the population in 2024 (Shofa & Muslim, 2024). The increasing trend of internet penetration has opened new economic opportunities, such as digital payment systems, remote working, e-commerce, cross-border transactions, and Artificial Intelligence (AI) tools.

However, the rise in internet penetration and digital platform adoption also presents significant challenges, as it increases the risk of scams spreading among the public. New technology such as AI are also being exploited for scams, such as replicating an individual's facial and audio features to deceive victims and illegally acquire financial gains. This crime, combined with existing online scams such as ransomware, phishing, and bogus schemes, poses an emerging threat, especially among vulnerable communities with lower levels of digital and financial literacy. In response to these threats, the Indonesian government has introduced several initiatives to combat scams, including the establishment of the Indonesia Anti-Scam Center. The Indonesian government has also launched its "National Strategy for Artificial Intelligence", providing a guideline to develop AI from 2020 to 2045. Despite existing laws on data protection, electronic information and transactions, and anti-scam strategies, there is room to strengthen policies and interventions to protect the public against online scams.

PATTERNS AND TRENDS

Types of Scams

Indonesia has historically been a victim of financial fraud cases, especially since the early 2000s due to the growing usage of mobile phones and the internet. The types of scams observed from 2008 to 2012 include distraction principle scams (advance fee/lottery scams, romance/relationship scams, job scams), time principle scams (intimidation scams via voice-and-text-based communication scams), kindness principle scams (charity scams via emails, websites, and forums), social compliance scams (business scams via SMS, email, forums), and unnoticeable scams (hacking, fake ATMs, phishing) (A.H. Kusomo, et al., 2017).

Unlike traditional scams that may rely on a single method, modern fraudsters frequently use a variety of channels, such as a combination of phone calls, emails, and text messages to boost their success rate. Moreover, creating a sense of urgency while impersonating law enforcement officers, bank officials, or family members is a key strategy in luring victims, as people tend to use less reasoning when making decisions under time pressure (A.H. Kusomo, et al., 2017).

Based on a 2013 State of The Internet report, Indonesia ranks second in the world for cybercrime cases, with romance scams being one of the most common, particularly among women. Typically, love scammers are foreign nationals working in groups with Indonesian accomplices to pretend as wealthy military officials, doctors, engineers, or businessmen. Fraudsters target individuals through social media,

[12] Researcher, Department of Economics, CSIS Indonesia
[13] Project Research Assistant, Safer Internet Lab, CSIS Indonesia
[14] Project Research Assistant, Safer Internet Lab, CSIS Indonesia

building an emotional relationship with them over a long time before manipulating them into sending money under the deception that they are helping a significant other. Victims tend to disregard warnings about the risks of being scammed because they are convinced that the relationship is real (Juditha, 2015).

In COVID-19, online scam cases have surged in tandem with the development and popularity of online platforms. Fraudsters offer victims illegal loans through SMS or WhatsApp from an unknown number. When they fail to pay on an agreed schedule, they can be charged a 10 percent interest daily, making the repayment of loans nearly impossible. Victims may fail to recognize they were being scammed, and are reluctant to report due to the fear of damaging their reputation. These fraudsters also threaten family members, using their personal data for their benefit, causing severe psychological consequences that may lead to suicide (Magfirah & Husna, 2022).

In 2021, the growth of online shopping platforms and digital payment systems has also opened new opportunities for e-commerce scams, where fraudsters act as fake sellers, fabricating the image of a trusted shop and luring victims to pay for goods that were never shipped (Saleh, 2022). In 2023, online gambling became a growing concern; they are frequently advertised through WhatsApp, offering promising bonuses to lure victims. Furthermore, perpetrators were found to quickly change their sites and mechanisms to avoid getting caught by law enforcement (Hasibuan, 2023).

In 2024, Indonesia ranks second in Southeast Asia with 85.908 phishing cases after Thailand, with nearly triple the number (Novianty, 2025). Previous studies show that most respondents were tricked into clicking seemingly credible organizations that promise rewards (33.8 percent) (Abisono, et al., 2022). Moreover, s.id, my.id, and biz.id were among the most used domains used by perpetrators to deceive victims, mainly targeting social networking sites (e.g., Facebook, Meta, Instagram), financial (e.g., Dana, BRI), and gaming platforms (e.g., Garena) (IDADX, 2024).

Moreover, scams are 3,5 times more likely to occur in piracy sites compared to mainstream sites, as most users are unaware of seemingly benign digital behaviors, such as watching sports streams on illicit websites. Out of the 90 advertisements identified through repeated page views, 52,22 percent of them were classified as high-risk entry points for online scams. Fake streaming platforms collect users' sensitive information by requiring them to create an account, and a user would likely encounter up to 20 cyber threats for every 100-page visit to these sites. For instance, links disguised as Adobe Flash installations were confirmed to have malicious capabilities, including potential attacks of malware, ransomware, credential theft, data exfiltration, and destructive attacks (Watters, 2024).

Targeted Victims

Individual targets

A study identified that Indonesia exhibits high levels of power distance and collectivism, which in turn contribute to a high level of materialism in the country. This culture drives the susceptibility of people to the temptation of quick financial gain schemes. In many cases (see Annex A), financially literate people are not immune to deceitful messages, as multiple psychological principles were used to lure them (Prabowo, 2024).

Despite that, ample evidence illustrates that fraudsters have been consistently impacting people from low socioeconomic and educational status. In 2022, Indonesia ranks 61st out of 100 countries based on the level of education and internet (Amanta, 2022), and low levels of financial literacy may contribute to the low awareness of loan scams (Magfirah & Husna, 2022). Fraudsters were also found to take advantage of this by offering them small amounts of cash in exchange for self-portraits of them with their national ID to be used for illegal activities (GrabDefence, 2022).

In terms of gender, in 2022, women accounted for 55 percent of online scam victims (BaliNews, 2025). In the following year, a qualitative study on romance scam victims in Indonesia revealed that physical attraction was the main determinant in pursuing the initial attraction online, as woman/s interpersonal success in the country is often associated with having a partner (Niman, et al., 2023). In terms of age group, financial scams appeal to younger and older individuals. Older individuals, especially those who live alone, are targeted due to their loneliness and less familiarity with online platforms (A.H. Kusomo, et al., 2017). Younger individuals, on the other hand, were targeted due to their susceptibility to indulge in addictive and trendy schemes.

The rise of online gambling in Indonesia has affected both the younger and the older generation. In 2024, approximately 4 million Indonesians engage in online gambling sites. Although the majority of players are between 31 and 50 years old (40 percent) and above 50 years old (33 percent), there is a small portion of underage players who are often neglected by their parents. Furthermore, about 80 percent of players come from the middle to lower socioeconomic class. From 2017 to 2022, online gambling transactions totaled up to a turnover value of Rp 190 trillion. Online gambling players in Indonesia are dominated by the West Java region (~535,644 players with transactions of Rp 3,8 trillion), followed by DKI Jakarta (~238,568 players with transactions of Rp 2.3 trillion) and Central Java (~201,963 players and transactions of Rp 1.3 trillion). Increased financial losses have led to massive debts, triggering them to commit criminal acts such as theft, fraud, and violence to earn money (Junaedi, 2024).

With having a simpler verification steps in registering digital financial services, the Indonesian Financial Transaction Reports and Analysis Centre (PPATK) found that more than 24.000 children aged 10 to 18 years were victims of financial scams, with transaction values of more than Rp 127 billion. Security features in digital wallets are becoming a double-edged sword, where user privacy measures were exploited to hide the identities and locations of fraudsters, easing cross-border transfers and untraceable payments. Fraudsters groom children to provide inappropriate content or services that are paid in virtual currencies, such as bitcoin, to conceal the traces of illegal funds (Yulia and Sofian, 2024; Febriansyah, et al., 2024; Fransisco, et al., 2024).

Furthermore, it is worth noting that Indonesia is not only a victim of online scams, but also a victim of human trafficking. The Director of Protection of Indonesian Citizens and Legal Entities from the Ministry of Foreign Affairs confirmed that as of February 2025, there have been 6.800 cases of Indonesian citizens involved in online scams since 2020, and the number will most likely continue to increase (AntaraNews, 2025). Syndicates attract victims on social media by advertising jobs (e.g., marketing, human resources, translation, finance, casinos, hotels, and IT) with simple and easy job requirements, false promises of high salaries, bonuses, free accommodations, logistics, and the opportunity to work overseas. However, some were also recruited by close peers, such as family members, friends, and neighbors (IOIM, 2023; LSCW, 2024).

Individuals who are in their twenties, have completed secondary education, and are bilingual were found trafficked to Cambodia, Myanmar, the Philippines, and Thailand as online scam operators (IOIM, 2024). Additionally, 53 Indonesians who were trafficked in Cambodia were recruited by Indonesian nationals, thus increasing trust during recruitment (UNODC, 2023). Although most victims were deceived and violently coerced into working in scam compounds (United Nations, 2025), some would proceed nonetheless due to attractive salaries and potential commissions for good performance. Challenges for law enforcement persist, as it may be that some people do not want or need to be rescued (UNODC, 2023).

Organizational/business targets

Fraudsters have primarily targeted financial institutions, with credit card thefts causing a loss of more than USD 100 million in 2008 alone (see Annex B). With the rising adoption of technology, from 2012 to 2015, the police have caught more than 497 suspects of cyberattacks, including financial scams, and more than half of them were foreign citizens. These fraudsters sent messages that appeared closely similar to a real token message from the bank, and security loopholes were reported in large e-commerce sites like Bukalapak, Tokopedia, and Sribu. Aside from financial motives, government and corporate websites were also hacked for political motives. Despite the major losses incurred from online scams in Indonesia, almost all cases were kept confidential with very limited public disclosure (Edy et al., 2017).

Emerging Trends

The use of Artificial Intelligence for online scams in Indonesia has only started to emerge in 2023, utilizing mainly deepfake and voice cloning technology to steal identities and deceive victims into believing they are someone trusted, and creating fake scenarios to trick them into transferring money. Most of the losses incurred were derived from deepfake videos, where faces of well-known public figures were stolen to verbally promote malicious sites or counterfeit programs, such as claiming they provide promising returns from online gambling or are distributing financial aid. These videos often have phone numbers attached for victims to contact before they are asked to transfer money in advance, usually with the promise of it being returned or as an administration fee, only to realize they will be sent to the fraudsters' accounts. On the other hand, news articles on voice cloning scams show that rechecking the person whose identity is being stolen effectively detects scams and avoids further damage (see Annex C on deepfakes cases in Indonesia).

POLICY ASSESSMENT AND ADDRESSING SCAMS

National Initiatives and Policies

Electronic Information and Transactions Law

The primary legal framework regulating the digital and cyber domain in Indonesia is the Electronic Information and Transactions (EIT) Law. Indonesia's EIT Law aims to regulate a clean, safe, ethical, productive, and just digital landscape to protect the public from misusing electronic information, documents, and transactions, and any cyberthreats. The table below outlines key provisions of the EIT Law that address cyberthreats and misinformation.

Table 4.1 Summary of Penalties Related to Online Scams in the Policy Document

Crime	Details	Penalties	Article
Online gambling	Creating, distributing, transmitting, and/or making electronic information and/or records related to gambling	Maximum of 10 years in prison or Rp 10 billion fine	45 (3) in (1/2024)
Disinformation	Distributing and/or transmitting electronic information and/or records with false information and causing	Maximum 6 years in prison and/or Rp 1 billion fine	45 A (1) in (1/2024)

	material loss		
Data theft	Unauthorized or unlawfully alters, adds, reduces, transmits, tampers with, deletes, moves, or hides electronic information and/or electronic records of other persons or the public.	Maximum 8 years in prison and/or Rp 2 billion fine	32 (1), 48 (1) in (11/ 2008)
Data theft	Unauthorized or unlawful moves or transfers of electronic information and/or electronic records to electronic systems of unauthorized persons.	Maximum 9 years in prison and/or Rp 3 billion fine	32 (2), 48 (2) in (11/ 2008)
Data theft	Compromising confidential information for public access	Maximum 10 years in prison and/or Rp 5 billion fine	32 (3), 48 (3) in (11/ 2008)
Malware, Ransomware, Trojan viruses	Unauthorized or unlawful acts resulting in faults in electronic systems, and/or resulting in them working improperly	Maximum 10 years in prison and/or Rp 10 billion fine	33, 49 in (11/ 2008)
Possession, distribution, or sale of tools	Unauthorized or unlawfully produces, sells, causes to be used, imports, distributes, provides, or owns hardware, software, passwords, or access codes for cybercrime	Maximum 10 years in prison and/or Rp 10 billion fine	34, 50 in (11/ 2008)
Phishing, identity theft, deepfake, voice cloning	Unauthorized or unlawful manipulation, creation, alteration, deletion, and tampering of electronic information and/or records with the intent that such seem authentic	Maximum 12 years in prison and/or Rp 12 billion fine	35, 51 (1) in (11/ 2008)

Furthermore, as a derivative regulation under the Electronic Information and Transactions Law, the Ministry of Communications and Digital Regulation (KOMINFO) 5/2020 requires electronic system providers in the private sector to ensure regulatory compliance in digital platforms. Electronic system providers may include e-commerce, financial services, digital content distribution services, communication services, search engines, and personal data processing services. Although the document only explicitly states terrorism, child pornography, and content that causes public distress as prohibited electronic content, online scams may fall under the last category.

Table 4.2. Summary of private-sector obligations in digital platforms based on the policy document

Obligations	Article
User-generated content must have governance regarding electronic information and/or documents, and provide reporting platforms	1 (7)
Electronic systems must not contain and spread prohibited information and/or documents	9 (3)
When instructed, electronic system operators must take down prohibited electronic information and/or documents at the latest 1 x 24 hours after a warrant is issued	11c
In the case that electronic system operators fail to take down prohibited electronic information and/or documents, the Ministry will take down or order Internet Service Providers to take down the electronic system (access blocking)	15 (7)
Electronic system operators who fail to take down prohibited electronic information and/or documents will receive an administrative fine according to the provisions of laws and regulations regarding non-tax state revenues	15 (10)
Electronic system operators are required to provide access to electronic systems and/or data to the ministries or institutions for supervision under specific laws and regulations	21
Electronic system operators established in foreign countries or are permanently domiciled in foreign countries but provide services, conduct businesses, and offer services in Indonesia must also be registered under the law.	4

Anti-Fraud Measures by The Financial Services Authority

The financial services authority (OJK) has outlined four main pillars for Financial Institutions (FIs) to establish anti-fraud strategies in POJK 12/2024, including online scams (e.g., online gambling, fictional online investments, online prostitution, and other crimes with financial motives). Based on the document, FIs may include banks, securities, insurance, brokerage companies, pension funds, venture capital, microfinance institutions, pawnbrokers, and other regulated financial entities, regardless of whether they are operated under conventional or Sharia principles. Furthermore, they must establish a working unit or function to manage anti-fraud strategies, where heads or officials in charge must have a certification and experience in anti-fraud, and/or adequate experience in relevant fields.

Table 4.3. Summary of the four pillars of anti-fraud strategy for Financial Institutions

Pillars	Strategy
Prevention	• Raise awareness of employees and stakeholders through training sessions, workshops, and internal policies • Actively assess operations most vulnerable to fraud • Impose strict hiring policies to filter trusted employees, and reward employees who adhere to anti-fraud measures • Educate customers on fraud patterns using various media (e.g., brochures, campaigns, posters)
Detection	• Ensure a secure and anonymous whistleblowing mechanism • Conduct random inspections, especially in high-risk units • Actively track transactions and operational activities • Establish a clear procedure for customers and employees to report fraud, and the steps to respond to them.
Enforcement	• Collect solid evidence for fraud cases (e.g., forensic accounting, digital track records, interviews) • Comprehensively report fraud cases to OJK based on their guidelines • Impose clear, fair, and strict sanctions on fraud perpetrators. • Ensure transparency during case investigations.
Assessment	• Maintain an updated database of fraud incidents. • Actively review patterns of past fraud cases and identify room to improve anti-fraud strategies. • Ensure all employees are updated with the latest anti-fraud strategy.

To protect the public from the increasing spread of scams in Indonesia, the Financial Services Authority (OJK) established a coordination body called SATGAS PASTI. This task force consists of multiple sectors, including the Ministry of Investment, the Ministry of Communication and Digital Affairs, the Ministry of Cooperatives, the National Police, the State Intelligence Agency (BIN), the National Cyber and Crypto Agency (BSSN), and others. The formation of this multi-sectoral task force is essential to address coordination challenges in combating scams, given the wide range of fraudulent activities—spanning investment scams, institutional fraud, and trade-related deception. Additionally, scams are not always a purely domestic issue; many originate from or are linked to foreign entities. To enhance its effectiveness, the task force established the Indonesia Anti-Scam Center (IASC), providing a platform for the public to easily report scam activities.

Data Protection Law

Indonesia has enacted the Personal Data Protection Law (UU PDP) under Law No. 27 of 2022, which establishes the principle that personal data is a fundamental right that must be safeguarded and protected. Personal data plays a crucial role in protecting the public from the threat of scams. The UU PDP sets standards and guidelines for all institutions—both private and governmental—that collect and process personal data. It aims to ensure that data is securely stored and handled to prevent misuse. While Indonesia has a strong legal foundation for data protection, as of the time this report was written, no specific institution has been designated to oversee its implementation and enforcement.

Other Initiatives

Private sector initiatives

Various sectors, including online platforms, e-commerce, and financial institutions have undertaken multiple initiatives to strengthen network security against the growing threat of online fraud and scams (see Annex D). Besides, private companies are also started to utilizing AI technology to upgrade their security measures, such as:

- Real-time fraud detection systems
- Machine learning to analyze transaction patterns, detect deepfakes, and prevent synthetic identity fraud.
- Enhanced security layers and Know Your Customer (KYC) processes through two-factor authentication (2FA) and biometric verification to prevent account takeovers and unauthorized transactions.
- New merchants undergo risk assessments before being allowed to list products, reducing scam storefronts.
- Automated scam reporting system: Users will soon be able to report scams directly via an integrated industry-wide portal.
- Deploying deepfake detection algorithms to flag suspicious videos and voices.

Multi-stakeholder initiatives

The cross-sectoral nature of online scams demands coordinated efforts across various stakeholders to enhance the effectiveness of prevention and response strategies. Addressing this issue in silos – whether by individual organizations or sectors – will limit the effectiveness of mitigation and case handling efforts. Insights from our expert survey indicate that there have been collaborative initiatives between the government and private sector, one example being the establishment of the Task Force for Eradication of Illegal Financial Activities *(Satgas PASTI)*.

On February 11th, 2025, the financial authority (OJK), the Task Force, and industry players (e.g., payment service providers, e-commerce, and other relevant parties) launched the Indonesia Anti Scam Centre (IASC) to respond to fraud reports according to applicable provisions. The public can report fraud incidents by contacting OJK's customer service at 157 or filling in a form at iasc.ojk.or.id with personal data (national ID, driver's license), proof of bank account ownership, chronological order of the incident, and proof of the transaction occurring.

Once a report is logged, IASC members will promptly conduct verifications, block fraudulent transactions, identify perpetrators, and coordinate legal actions with law enforcement. Victims can track their report progress via the IASC system or the customer services of the financial services platform affected. The retrieval of lost funds can be attempted, though it may take time to coordinate between banks and financial services under the following conditions:

1. The recipient's account still has remaining funds, and transactions were verified originating from the victim's account.
2. In the case of multiple victims, refunds will be prioritized based on the order of refund requests received, the availability of remaining funds in the victim's account, and a mutual agreement among the victims.
3. If the victims fail to reach an agreement, refunds will be distributed based on a final court ruling.
4. A refund may not be processed if the recipient's account is blocked or seized by authorities.

Other initiatives include a UNODC webinar on the Digital Financial Threat Landscape and Law Enforcement in Indonesia. This webinar was funded by the Ministry of Justice of the Government of the Republic of Korea, inviting discussions with PPATK, the Director of Criminal Justice at Optima, and the Crime Prevention and Criminal Justice Officer at the Terrorism Prevention Branch of the UN Office on Drugs and Crime. The webinar highlighted the challenges in investigating financial scams and the urgency for increased information security, stakeholder mapping, and cross-border collaborations in cybercrime investigations (UNODC, 2022).

In addition, there is also an initiative to have sharing databases as to improve the warning systems from the potential online scams treats. This sharing database systems include:

- Cross-platform fraud intelligence sharing: E-commerce companies share scammer lists to prevent repeat offenders from migrating between platforms.
- Merchant blacklist system: Shared database of fraudulent merchants, ensuring they cannot reopen accounts easily.
- International cooperation: Some platforms work with Interpol and ASEAN cybersecurity bodies to track cross-border fraud networks.

CHALLENGES IN ADDRESSING SCAMS

Despite various initiatives undertaken by different stakeholders, Indonesia continues to face significant challenges in addressing scams. These challenges are not limited to the government alone but also extend to other parties, as combating scams requires strong inter-agency cooperation to enhance public protection.

Government Institutions

Challenges in Consumer Protection and Personal Data Framework

One of the key challenges in consumer protection efforts in Indonesia is the fragmented nature of regulations across different sectors, coupled with the outdated Consumer Protection Law. Indonesia's Consumer Protection Law was enacted in 1999, making it ill-equipped to address emerging challenges in the digital era. Although a new Consumer Protection Law has been included in the national legislative priority program (Prolegnas), Indonesia has yet to enact an updated version.

Another challenge in safeguarding consumers and the public is the personal data protection framework. While Indonesia has already passed the Personal Data Protection Law (UU PDP), the necessary implementing regulations have yet to be issued. These regulations are essential for providing clearer guidelines, enabling both government and private institutions to comply effectively with the UU PDP. Additionally, an independent oversight body has not yet been established to monitor data misuse. Without a dedicated regulatory body, ensuring personal data security and strengthening data infrastructure and governance will be significantly more challenging.

As technology continues to advance—especially with the increasing use of generative AI in scams—the role of data infrastructure and governance becomes even more critical. Without a robust data governance framework, scammers could exploit personal data for highly targeted scams. In the worst-case scenario, deepfake technology could be used to mimic family members, making individuals even more vulnerable than before.

Weak Enforcement Measures

As previously discussed, Indonesia currently lacks a single authoritative institution responsible for handling scam-related issues. Institutional silos remain a major challenge, as each institution has its own mechanisms for addressing scams. This fragmented approach creates a lack of clarity for the public on where and how to report scams, ultimately discouraging victims from coming forward.

Another issue is the complexity of the reporting system, which is often seen as burdensome by the public. Many victims find the reporting process non-straightforward and difficult to navigate. Furthermore, expert interviews indicate that only a small number of reports receive official responses. The lack of response could stem from regulatory bottlenecks or incomplete documentation from the complainants—an issue that needs further investigation. To improve scam prevention efforts, enhancing the reporting mechanism and providing assistance to complainants could help eliminate documentation issues on the victims' end. This, in turn, would allow regulators to respond more effectively to scam-related complaints.

Private Sector

The private sector's challenges in combating online scams in Indonesia include inconsistent policies across jurisdictions, privacy laws restricting collaboration between platforms and authorities, and incomplete scam data that may reduce response effectiveness. In detail, past studies have highlighted challenges in the following platforms:

a) **Digital banking** systems may impose extensive identity verification measures that users find complicated. As a result, a portion of Indonesians have stopped or reduced the use of credit cards (32 percent) and personal bank accounts (27 percent) due to finding identity checks too difficult and time-consuming. However, more than half surveyed had a strong preference and belief that fingerprint and face scans provide excellent security. In Asia Pacific, 60 percent would only answer up to 10 questions before abandoning the platform and seeking alternatives. Additionally, 23 percent of the Indonesian sample believes there are circumstances when falsifying personal information for digital loan applications is acceptable, while 11 percent believe it is a normal practice (FICO Consumer Survey, 2023).

b) **E-wallets** do not require strict identification, allowing fraudsters to create alternate identities by relying on phone numbers to register and cash out without a bank account. The platform can only see transaction amounts and parties, with little information on the purpose and intent of the payments, enabling them to transact without leaving a significant trace. To avoid suspicion, funds are transferred in small amounts, making detection in transaction monitoring systems increasingly challenging (Fransisco, et al., 2024; Yulia and Sofian, 2024).

c) **E-commerce** platforms may offer "Cash on Delivery (COD)" methods, which can be exploited by buyers who refuse to pay and sellers who ship incorrect goods with limited protection for both parties. Due to gaps in labor protection, couriers are also vulnerable to e-commerce scams (Fadillah, et al., 2023). Additionally, transactions conducted outside official platforms—without a third-party intermediary—such as social commerce are often exploited by scammers, leaving victims more vulnerable. Despite existing regulations regarding online scams, there are still regulations that do not specifically explain fraud involving e-commerce, and the collection of digital evidence remains a challenge (Widhaningroem and Widowati, 2024).

d) **Social media** platforms are vulnerable to click-farming operations, where informal workers provide services to inflate engagement metrics (e.g., likes, views, followers) by generating fake accounts using bots or software that run automated scripts, harvesting real accounts from

exchanged or stolen login credentials via websites offering free followers, and selling these services to online shops or public figures, such as influencers and politicians. A mixture of fake and real accounts continues to challenge detection algorithms, and the entry barrier to becoming a perpetrator is low because websites may offer simple procedures that do not require programming skills (Lindquist, 2018).

e) **The telecommunications industry's role** needs to be strengthened in efforts to combat scams in Indonesia. This is because the primary channels for scam distribution are messaging apps and phone calls. The linkage between phone numbers and national identity numbers (NIK) presents an opportunity to more effectively address scams, not only within the telecommunications sector but also in digital services that require phone numbers for registration, such as e-commerce, digital financial services, and others.

   One key initiative that should be encouraged is greater collaboration between the telecommunications industry and other sectors. While telecom companies have already introduced the Open Gateway initiative, its benefits have yet to be fully utilized by other industries that scammers exploit for fraud. In addition, the implementation of facial recognition technology for SIM card registration should be welcomed as a positive step in scam prevention. However, the potential misuse of facial recognition remains a challenge, especially with the rise of deepfake technology, which could further complicate security measures in the future.

Public

One of the main challenges in preventing scams in Indonesia is that increased internet access has not been accompanied by improvements in digital skills and literacy. Overall, Indonesia's digital literacy rate remains low, while the awareness related to digital security is even lower. The Indonesian Digital Society Index 2024 reports that out of the four pillars adapted from the G20 Toolkit for Measuring Digital Skills and Digital Literacy stated that:

- Indonesia scored the highest in the digital skills pillar (58,25 percent), while it scored the lowest on the empowerment pillar (25,66 percent), which shows that the increase in digital technology competencies remains insufficient to support productive economic activities.
- Less than 50 percent of the sample have a habit of ensuring the credibility of sources when engaging with digital information.
- Less than 50 percent of the sample stores backup data, use two-step verifications, and understand security threats in digital tools.
- Low adoption of digital upskilling, as only 5 percent have followed online courses, and only 2 percent have been paid instructors in an online course.
- Low usage of digital financial tools, such as Internet/mobile banking (40 percent), e-wallet (38 percent), online investments (4 percent), online loans (5 percent), and online lending (1 percent)

Another major challenge in combating scams in Indonesia is the low level of financial literacy among the public. Financial literacy can be defined as a set of skills that enables individuals to comprehend and manage their finances, including planning future finances, managing risks, and actively participating in financial markets (Arifin, et al., 2024). Without sufficient knowledge of digital financial products, people are more susceptible to scams, particularly those promising unrealistic investment returns. Based on the National Survey on Financial Literacy and Inclusion 2024 by OJK:

- The financial literacy index has increased from 21,854 percent in 2013 to 65,43 percent in 2024.

- Despite that, disparities persist between the population in urban (69,71 percent) and rural areas (59,25 percent).
- The lowest scores were in the 15-17 age category (51,70 percent) and the 51- 79 age category (52,51 percent).
- In terms of gender, females (66,75 percent) ranked higher in financial literacy than males (64,14).
- Based on occupation, unemployed (42,18 percent), students/university students (56,42 percent), retired persons/military veterans (57,55 percent), farmers/gardeners/fishers, and occupations other than employees, professionals, and entrepreneurs are of concern.
- Education levels are also in parallel with financial literacy, as those who completed university had the highest financial literacy index (88.29 percent), and those who did not enroll in any formal education scored the lowest (51,53 percent).

Moreover, in terms of cybersecurity enforcement, 30 percent of problems in Indonesia's cybersecurity landscape are due to the shortage of cybersecurity experts (Saleh and Winata, 2023). Based on Marwi and Oskar (2023), users may be discouraged from reporting scam cases they have encountered due to the following reasons:

- Lack of understanding of reporting systems
- Unsatisfactory banking hotlines with high phone credit rates
- The feeling of embarrassment and the feeling that the problem could not be resolved. Victims with higher education backgrounds refused to report their cases due to embarrassment.
- The police claimed to update the victims on the case, but there were no further updates from them
- The emphasis on the burden of proof was placed on victims
- Despite having seen warnings of scams, they did not remember them when they encountered similar schemes

RECOMMENDATIONS

To address these challenges, a series of initiatives must be implemented in multiple stages. Ideally, the most effective way to combat scams is by enhancing public literacy and awareness of fraudulent activities. However, this is not an easy task and requires a long-term effort. Given the current low levels of digital literacy in Indonesia, regulators and other stakeholders must focus on proactive measures to tackle scams more effectively. The following are key strategies that should be pursued to bridge the existing policy gaps in Indonesia:

a) **A data sharing mechanism involving relevant stakeholders from the private and public sectors.** When a scam attack occurs on a digital platform, records of the perpetrator *(e.g., the pattern of attack, IP address, identities)* are documented in a collaborative database where other stakeholders, such as e-commerce, banks, the government, and financial technology companies can access. Thus, when a user who matches the identities of the perpetrator is identified in another digital platform, they would not be able to continue creating or using their accounts until further inspections. Although our expert survey reported that members of an e-commerce association have a shared merchant fraud database, a similar mechanism has yet to be implemented by other actors.

   Of course, the data sharing initiative must be conducted in accordance with proper data governance principles, as mandated by the Personal Data Protection Law and robust international best practices. One of the main challenges in implementing data sharing is the

existence of confidentiality regulations, particularly in sectors like banking, which make data exchange difficult. Additionally, the absence of implementing regulations for the UU PDP creates further uncertainty regarding the extent to which institutions can share data and whether certain exceptions may be allowed for specific purposes. To effectively combat scams in Indonesia, it is essential to establish a legal framework for data sharing that is specifically designed for fraud prevention. This framework must be built on strong governance principles while prioritizing personal data protection.

b) **The use of AI for scams prevention.** The use of new technologies, such as AI, should also be leveraged to address scams. AI has the potential to analyze suspicious data from scammer behavior across various platforms—whether in messaging apps, phone calls, or digital platforms—enabling early detection of potential scams. This would allow the public to be notified of possible scam threats in real time.

   However, AI and technology are not silver bullet solutions. First, our understanding of AI is still limited, making it challenging to fully map the opportunities and risks associated with its use. Despite the need to encourage the use of AI for this purpose, fundamental principles such as transparency, fairness, accountability, and others must remain a cornerstone of its application. Second, it is also important to first assess existing horizontal regulations and laws whether the current framework sufficiently addresses emerging AI risks. Leveraging existing regulatory and legal frameworks within specific sectors provides an adaptable foundation for AI governance. This approach supports ethical AI development while enabling targeted, effective responses to evolving threats like deepfakes and sophisticated scams.

c) **A single institution designated to surveil, document, and respond to online scam cases.** The existing policy landscape illustrates an unclear division of roles and responsibilities in addressing scams between the government, such as KOMINFO, and the financial regulator, OJK. In simple terms, Indonesia currently has various methods for reporting scams, for instance the one that is under KOMDIGI and OJK. However, the absence of a single designated institution and a clear mechanism creates confusion among the public. Additionally, there are challenges related to the complex and time-consuming reporting procedures. Therefore, there is a need for one unified reporting mechanism that is simple, easy to remember, accessible, and user-friendly, to facilitate the public in reporting scam activities.

d) **Targeting digital literacy interventions to include scam awareness, identification, and reporting.** Established efforts to promote digital literacy lack depth for the public to be equipped with the knowledge and confidence in identifying scams. Other than public awareness campaigns on print and online media, programs leaned more towards onboarding sessions for users to operate digital platforms, such as training merchants to utilize e-commerce applications. To ensure program effectiveness, interventions should prioritize vulnerable populations, such as the older generation and those from a lower socioeconomic background. Furthermore, scam awareness can also be integrated into the formal educational curriculum at all levels, as has been done by OJK to enhance digital literacy.

e) **Expanding multi-level international cooperation.** As online scams operate in the digital space, cases of foreign syndicates and cross-country victims demand multinational cooperation. Collaboration between anti-scam initiatives across countries needs to be expanded, one of which can be achieved by establishing common standards and guidelines for addressing cross-border scams. A unified framework and set of standards are essential, covering areas such as

prevention efforts, response to the spread of scams, victim protection, and more, especially considering that the leading institutions overseeing anti-scam initiatives vary by country.

Multilateral cooperation (not just bilateral) is critical because domestic efforts alone will never be effective, particularly since the spread of scams via the internet transcends national jurisdictions. This cooperation should also extend beyond just tackling the spread of scams and victim protection to include the growing concern of human trafficking. It is hoped that multilateral collaboration will increase the effectiveness of governmental efforts in combating scams.

ANNEXES

Annex A. Table of a sample online scam cases targeting individuals in Indonesia

Year	Platform	Scam type	Details	Socioeconomic implications
2016	OLX, Kaskus, Bukalapak, Tokopedia (Sasongko, 2016)	E-commerce scam	Fraudsters created fake online stores, received money from customers, and deleted purchasing history	93 victims were reported, with a loss of Rp 10,1 billion
2016	SMS (Pinrang & Syamsuddin, 2016)	Prize scam	Fraudsters sent randomized text messages to 5.000 phone numbers claiming they won millions of rupiah. Prizes can be redeemed after a specific transfer amount.	For every Rp 10.000 of transactions, one fraudster received Rp 3.000. 3 laptops, 6 cellphones, hundreds of phone cards, and Rp 65 million were confiscated.
2016	Phone call (Amelia, 2016)	Lottery scam	Fraudsters from Jakarta, Surabaya, and China called victims claiming they won money from a lottery. Prizes can be redeemed if they paid Rp 2,6 billion	52 satellite phones, 6 handphones, and 8 laptops were confiscated
2020	Instagram (Luxiana, 2020)	Shopping scam	Four 15-16-year-olds claimed to sell rare and branded items on several Instagram accounts, but have never sent anything after receiving money from customers.	Dozens of victims with a loss of over Rp 100 million.
2022	Online loan platform (BBC, 2022)	Investment scam	A bogus investment scheme promises a 10 percent return on investment every month. The return was only paid on the first month.	331 victims with a loss of Rp 2 million to Rp 10 - Rp 19 million each, and an estimated total loss of Rp 2,1 billion.
2022	Instagram (Tempo, 2022)	Shopping and impersonation scam	A fraudster impersonating Indonesia's Customs Office contacted victims claiming an online purchase was illegal, forging documents, and demanding payments.	A total loss of Rp 11.5 million for one victim and Rp 8.1 million for another.
2024	E-commerce (Syafaruddin, 2024)	Dropship-ping scam	A fraudster posted a dropshipping ad, luring the victim to pay for a 'warehouse'. When there was an expensive order, they offered her a 40% loan. The victim was tricked into believing the business was profitable, and she was blocked from withdrawing her money from the fake e-commerce site.	A total loss of Rp 115 million for the victim.

2024	INDODAX, Cryptocurrency platform (Greig, 2024)	Cryptocurrency fraud	Fraud under the name of INDODAX; sending out refund invitations or personal requests in the platform, which has 5 million users.	At least US $230 million worth of cryptocurrency was lost
2024	Instagram (Setiawati, 2024)	Romance and impersonation scam	Fraudsters approached victims on Instagram; in one case posed as an oil and gas engineer in Papua, claiming their salary was withheld and had a heart attack with fake medical bills and funeral costs. Victims were contacted by fake 'friends' to continue the fraud.	The largest loss was recorded by a single parent of over 50 years old with Rp 600 million, while others lost millions to hundreds of millions of rupiah.
2024	WhatsApp (CNN, 2024)	Malware	Fraudsters sent an application scam (.apk file) disguised as a letter from the regional police via WhatsApp. Once clicked, hackers can access the victim's SMS, allowing them to steal OTP codes and drain bank accounts.	Undisclosed
2025	Dating application (The Jakarta Post, 2025)	Romance, investment, cryptocurrency scam	Gambir police arrested 20 suspects, headed by a Chinese national were suspected of creating fake identities to lure mostly women foreigners to invest in fraudulent cryptocurrency scams.	The operation's leaders earned Rp 7 million, while scammers earned Rp 5 million

Annex B. Table of a sample online scam cases targeting organizations in Indonesia

Year	Organization	Scam type	# of data affected	Socioeconomic implications
2013	Garuda Indonesia (Panji, 2013)	Data theft	<20 credit card records	Undisclosed
2013	PT. Bumi Resources Tbk (BUMI) (detikFinance, 2013)	Data breach (Trojan virus)	All data from 3 computers (1,5 terabytes)	Undisclosed
2015	Mandiri Bank (Heriyanto, 2015)	Malware	Undisclosed, one claimed losing up to Rp 13 million	Undisclosed
2017	Tiket.com (CNN, 2017)	Data theft	Undisclosed; 4 syndicates successfully hacked 400 other sites	Loss of Rp 1,9 billion
2020	Tokopedia (Mulia, 2020)	Data theft	91 milion user records	Sold for US$ 5.000 on the dark web
2020	Kreditplus (Clinten & Yusuf, 2020)	Data theft	890 thousand records for 78 MB of data	Sold for about Rp 50.000 on the dark web
2021	Indonesia's Healthcare and Social Security Agency (BPJS Kesehatan) (Mulia, 2021)	Data theft	279 million records	Undisclosed

2022	Bank Indonesia (CISO MAG, 2022)	Ransomware	Undisclosed; mitigation measures were undertaken and had no impact on critical data	Undisclosed
2023	Directorate General of Immigration (Hope, 2023)	Data theft	34 million passport records	Undisclosed
2023	Bank Syariah Indonesia (BSI) (The Cyber Express, 2023)	Ransomware	1.5 terabytes of data; over 15 million customers and employees	Data was released due to failure to meet the ransom demand of US$ 20 million
2024	Taxpayer Identification Numbers (NPWP) (Dinilhag, 2024)	Data theft	6 milion taxpayer records	Sold for US$ 10.000 on the dark web
2024	National Data Centre Ransomware Attack (PDNS) (Reuters, 2024)	Ransomware	Disrupted 160 government agencies, including immigration and airport operations	US$ 8 million to unlock encrypted data, but later apologized and decrypted data without payment
2024	General Elections Commission (KPU) (Paganini, 2024)	Data theft	252 million voter records	Sold for US$ 74.000 or 2 bitcoins

Annex C. Table of a sample online scams using AI technology in Indonesia

Year	Platform	Scam type	Details	Socioeconomic implications
2023	E-commerce (Expert Survey)	Deepfake	Fraudsters used deepfake images and videos to impersonate real sellers or executives	Financial losses: Chargebacks and refunds, revenue loss, additional operational costs for security measures, legal fees, and investigation Non-financial losses: Reputation damage, regulatory and compliance risks
2024	Various online platforms (CNN, 2024)	Deepfake	A deepfake clip of Indonesian public figures, such as Najwa Shihab, Raffi Ahmad, and Atta Halilintar promoting online gambling sites circulated the internet.	Undisclosed
2024	WhatsApp (Fernando, 2024)	Voice cloning	A victim claimed receiving a text and phone call from someone he thought was his close friend. He did not suspect anything as it sounded exactly like him. The fraudster offered cheap auction items (e.g., iPhone, Canon camera, electronics) for Rp 10 million, encouraging him that it could be sold for a more expensive price.	No financial losses were incurred; the victim rechecked with their friend and found their WhatsApp status warning that there have been scam incidents under the guise of their name.
2025	WhatsApp (France 24, 2025; Tempo, 2025)	Deepfake	A deepfake clip of the Indonesian president, Prabowo Subianto, uttering *"Who hasn't received aid from me? What are your needs right now?",* was circulated with a WhatsApp number attached. Victims were asked to transfer Rp 250 thousand to Rp 1 million as an "administrative fee" to redeem the 'aid'.	100 people were scammed from 20 provinces. A suspect pocketed Rp 65 million from the scam.
2025	Social media	Deepfake	A deepfake clip of Indonesian public figures, such as the vice president, Gibran Rakabuming Raka, and the	A loss of approximately Rp 30 million in the last 4 months.

	(CNN, 2025)		finance minister Sri Mulyani offering financial aid for citizens in need, with a phone number attached as a 'call center.' Victims were asked to fill in a registration form with an administration fee.	
2025	WhatsApp (Aprilia, 2025)	Voice cloning	A victim received a phone call from someone they thought was their friend, with the possibility of voice cloning.	No financial losses were incurred; the victim rechecked with their friend and found that the same incident had happened to three other people.
2025	Digital payment system (Expert Survey)	Not specified	A fraudster might have used AI to replicate identities	Fraudsters did not repay the financial damages, and the platform was reported to the police

Annex D. Table of a sample of private sector initiatives to prevent online scams

Sectors	Company	Initiative
Digital financial services	Dana	Dana joins the Global Anti-Scam Alliance (GASA) (2025): This collaboration gains Dana access to a global network of experts to enhance fraud prevention with AI-powered detection and cybersecurity innovation, among others.
E-commerce	Shopee	Shopee Guarantee (2024): A product can be labeled '100% Original' when it meets specific criteria; the label can only be declared by the Shopee team to protect users from shopping fraud.
E-commerce	Blibli	Partnership with Vesta (2023): Blibli hired Vesta, a US-based transaction guarantee platform that uses machine learning, AI, and global data to protect Blibli's card-not-present (CNP) transactions, eliminating the risk of chargebacks, and an enhanced fraud management system.
Bank	Bank Central Asia	Don't Know? Say No! (2023): Campaign to raise awareness of banking fraud, featuring a public figure in the entertainment industry with more than 8 million views on YouTube.
E-commerce	Tokopedia	Penalty System: Merchants that manipulate transactions, reviews, and products sold to enhance their shop's reputation, misuse promos for personal matters, the use bots can be given penalties, such as a permanent account/store closure, withdrawal of subsidies, or a deduction of balances from the Merchant.
E-commerce, ride-hailing, fintech	Grab, Ovo, Link Aja	GrabDefence: Provides access to a full suite of tools, technologies, and intelligence that are tried and tested in millions of data points for real-time protection against fraud and financial crime through 3 steps: 1. Provides a set of APIs for tools enabling data collection, data preparation, and tool integration. 2. Applies a combination of risk rules and machine learning algorithms at key checkpoints of the user's journey to instill protection. 3. Iteratively managing risks through continuous monitoring, experimentation, and fine-tuning.

E-commerce, ride-hailing, fintech	Go-jek	JARVIS (2018): A tool to automate manual data retrieval and basic analysis. What used to take a human analyst 30 minutes now takes 3 seconds to prevent fraud at scale and speed.
Telco industry	Telkomsel, Indosat, XL, Smartfren	Open Gateway initiatives to improve more robust authentication, detects SIM swap service, detect potential fraud from GPS manipulations.

REFERENCES

6,800 Indonesians involved in online fraud overseas: Govt. (2025, February 21). Antara News. https://en.antaranews.com/news/345845/6800-indonesians-involved-in-online-fraud-overseas-govt

Amanta, F. (2022, July 6). *Unpacking Indonesia's Digital Accessibility*. CIPS | Think Tank. https://www.cips-indonesia.org/post/opinion-unpacking-indonesia-s-digital-accessibility

ASEAN-Australia Counter Trafficking. (2024, May). *Human Trafficking & Forced Labour in Cambodia's Cyber-Scam Industry*. https://www.aseanact.org/wp-content/uploads/2024/05/202405-LSCW-Cyber-scams-and-HT-report-design-1.pdf

Assessing the Digital Financial Threat Landscape in Indonesia. (2022, April). https://www.unodc.org/roseap/en/what-we-do/anti-corruption/topics/2022/03-assessing-digital-financial-threat-landscape-indonesia.html

Badan Pengembangan Sumber Daya Manusia Komunikasi dan Digital. (2024). *Indeks Masyarakat Digital Indonesia 2024*. https://imdi.sdmdigital.id/publikasi/02122024_Buku%20IMDI_BAB%201-5_V6_compressed.pdf

Badan Pengkajian dan Penerapan Teknologi. (2020). *Strategi Nasional Kecerdasan Artifisial Indonesia 2020—2045*. https://ai-innovation.id/images/gallery/ebook/stranas-ka.pdf#page=1

Bank Syariah Indonesia Cyber Attack: LockBit Demands $20m Ransom. (2023, May 16). https://thecyberexpress.com/lockbit-bank-syariah-indonesia-cyber-attack/

Begini Cara Hacker Bobol Situs Tiket.com. (2017, March 31). CNN Indonesia. https://www.cnnindonesia.com/teknologi/20170331145137-185-204065/begini-cara-hacker-bobol-situs-tiketcom

CISOMAG. (2022, January 21). Bank Indonesia Suffers Ransomware Attack, Suspects Conti Involvement. *CISO MAG | Cyber Security Magazine*. https://cisomag.com/bank-indonesia-suffers-ransomware-attack-suspects-conti-involvement/

DANA Indonesia. (2025, February). *DANA Joins GASA*. https://www.linkedin.com/posts/dana-indonesia_dana-joins-gasa-activity-7307355028527140865-Z_pa/

Edy, S., Gunawan, W., & Wijanarko, B. D. (2017). Analysing the trends of cyber attacks: Case study in Indonesia during period 2013-Early 2017. *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, 1–6. https://doi.org/10.1109/INNOCIT.2017.8319146

Email Perusahaan Tambang Bakrie Kena Hack, Seluruh Data Dicuri. (2013, February 12). detikfinance. https://finance.detik.com/bursa-dan-valas/d-2168234/email-perusahaan-tambang-bakrie-kena-hack-seluruh-data-dicuri

Fadillah, T. (2023). *E-Commerce: A New Media that Creates New Disasters*. OSF. https://doi.org/10.31219/osf.io/hpb6w

FICO. (2023). *Consumer Survey 2023—Digital banking, customer preferences, and fraud controls*. Retrieved April 17, 2025, from https://www.fico.com/en/latest-thinking/ebook/consumer-survey-2023-digital-banking-customer-preferences-and-fraud-controls

Forrester. (2022). *Staying Ahead Of The Fight Against Fraud In Southeast Asia (SEA)*. https://assets.grab.com/wp-content/uploads/sites/4/2022/02/22102958/Forrester_Staying_Ahead_Of_The_Fight_Against_Fraud_SEA.pdf

Fraud, Scam and Fincrime Detection, Digital Risk Security Solutions | GrabDefence. (n.d.). Retrieved April 17, 2025, from https://defence.grab.com/

Greig, J. (2024, September 13). *Largest crypto exchange in Indonesia pledges to reimburse users after $22 million theft*. https://therecord.media/indodax-crypto-exchange-pledges-to-reimburse-after-theft?utm_source=chatgpt.com

Harwanto, F., Febriansyah, A., & Irwantika, N. (2024). *Cryptocurrency, Crime, And Children: Unveiling The Dark Side of Financial Technology in Child Sexual Exploitation*. 29–35. https://doi.org/10.2991/978-2-38476-325-2_4

Hasibuan, E. S. (2023). The Police are Indecisive: Online Gambling is Rising. Facts about the Eradication of Online Gambling in the Field. *Journal of Social Research*, *2*(10), 3365–3370. https://doi.org/10.55324/josr.v2i10.1405

Heriyanto, T. (n.d.). *Selain BCA, Nasabah Mandiri Juga Kena Malware Pencuri Uang*. teknologi. Retrieved April 17, 2025, from https://www.cnnindonesia.com/teknologi/20150306104201-185-37163/selain-bca-nasabah-mandiri-juga-kena-malware-pencuri-uang

Herman, A. D. (n.d.). *6 Million Taxpayer IDs, Including President's, Allegedly Leaked and Sold for $10,000*. Jakarta Globe. Retrieved April 17, 2025, from https://jakartaglobe.id/news/6-million-taxpayer-ids-including-presidents-allegedly-leaked-and-sold-for-10000

Hope, A. (2023, July 13). 34 million Indonesian Passports Exposed in a Massive Immigration Directorate Data Breach. *CPO Magazine*. https://www.cpomagazine.com/cyber-security/34-million-indonesian-passports-exposed-in-a-massive-immigration-directorate-data-breach/

Indonesia Anti-Phishing Data Exchange. (2024). *Laporan Aktivitas Abuse Domain .ID*. https://api.idadx.id/documents/uploads/1724725529_Laporan%20Q2%202024.pdf.pdf

Indonesian ecommerce platform Blibli hires Vesta for payments security. (2023, January 24). https://thepaypers.com/digital-identity-security-online-fraud/indonesian-ecommerce-platform-blibli-hires-vesta-for-payments-security--1259990

IOM UN Migration. (n.d.). *Information on Forced Labor and Trafficking in Persons (TIP)—Indicated Cases in Online Scamming Industry Overseas*. https://indonesia.iom.int/sites/g/files/tmzbdl1491/files/documents/2023-08/infosheet-online-scams-english.pdf

IOM UN Migration. (2024, February). *IOM's Regional Situation Report on Trafficking in Persons into Forced Criminality in Online Scamming Centres in Southeast Asia.* https://roasiapacific.iom.int/sites/g/files/tmzbdl671/files/documents/2024-02/iom-southeast-asia-trafficking-for-forced-criminality-update_december-2023.pdf

Juditha, C. (2015). *Pola Komunikasi Dalam Cybercrime (Kasus Love Scams)*. *6*(2). https://media.neliti.com/media/publications/122582-ID-none.pdf

Junaedi, J. (2024). England is the Largest Center for Online Gambling Activity in the World, Versus Indonesia is Exposed to Online Gambling Emergency Stage Five. *International Journal of Law, Crime and Justice*, *1*(3), 100–114. https://doi.org/10.62951/ijlcj.v1i3.134

Khan, M. A. (2024). Understanding the Impact of Artificial Intelligence (AI) on Traditional Businesses in Indonesia. *Journal of Management Studies and Development*, *3*(02), Article 02. https://doi.org/10.56741/jmsd.v3i02.584

Kristanto, K., Ismail, K., Fransisco, F., & Ronaldi, R. (2024). *Criminal Liability of Child Sexual Exploitation Perpetrators Using Social Engineering Techniques Through Digital Wallets in Indonesia*. 20–28. https://doi.org/10.2991/978-2-38476-325-2_3

Kurni, N., Rahayu, Wendratama, E., Monggilo, Z., Damayanti, A., Angendari, D. A., Abisono, F., Shafira, I., & Desmalinda. (2022). *Penipuan Digital di Indonesia*. https://cfds.fisipol.ugm.ac.id/wp-content/uploads/sites/1423/2022/08/PDF-Monograf-Penipuan-Digital-di-Indonesia-Modus-Medium-dan-Rekomendasi.pdf

Law of the Republic of Indonesia Number 11 of 2008 Concerning Electronic Information and Transactions. https://www.icnl.org/wp-content/uploads/Indonesia_elec.pdf

Lindquist, J. (2018). Illicit economies of the internet: Click farming in Indonesia and beyond. *Made in China Journal*, *3*(4), 88–91. https://doi.org/10.3316/informit.035564674568222

Luxiana, K. M. (n.d.). *Puluhan Orang Korban Penipuan Online Sindikat Bocah SMP Rugi Rp 100 Juta*. detiknews. Retrieved April 17, 2025, from https://news.detik.com/berita/d-5178941/puluhan-orang-korban-penipuan-online-sindikat-bocah-smp-rugi-rp-100-juta

Maghfirah, F., & Husna, F. (2022). CYBER CRIME AND PRIVACY RIGHT VIOLATION CASES OF ONLINE LOANS IN INDONESIA. *PROCEEDINGS: Dirundeng International Conference on Islamic Studies*, 1–18. https://doi.org/10.47498/dicis.v1i1.1009

Marwi, H., & Oskar, I. (2023). Analysis Of Increasing Types Of Online Fraud And Level Of Public Awareness In Indonesia. *Journal of Embedded Systems, Security and Intelligent Systems*, 70–84. https://doi.org/10.59562/jessi.v4i2.722

Media, K. C. (2013, November 21). *Situs Webnya Diserang, Ini Penjelasan Garuda Indonesia*. KOMPAS.com. https://tekno.kompas.com/read/xml/2013/11/21/1724512/Situs.Webnya.Diserang.Ini.Penjelasan.Garuda.Indonesia

Media, K. C. (2016, February 13). *Tujuh Pelaku Penipuan via SMS Dibekuk, Salah Satunya Pegawai Bank*. KOMPAS.com. https://regional.kompas.com/read/xml/2016/02/13/11585501/Tujuh.Pelaku.Penipuan.via.SMS.Dibekuk.Salah.Satunya.Pegawai.Bank

Media, K. C. (2020, August 4). *Data Ratusan Ribu Nasabah Kredit Plus Diduga Bocor dan Dijual di Internet*. KOMPAS.com. https://tekno.kompas.com/read/2020/08/04/07150007/data-ratusan-ribu-nasabah-kredit-plus-diduga-bocor-dan-dijual-di-internet

Mulia, K. (2020, May 6). *What can we learn from Tokopedia's alleged 91-million data leak?* KrASIA. https://kr-asia.com/what-can-we-learn-from-tokopedias-alleged-91-million-data-leak

Mulia, K. (2021, June 15). *Indonesians' personal information is up for sale. Who's buying?* KrASIA. https://kr-asia.com/indonesians-personal-information-is-up-for-sale-whos-buying

Naibaho, R. (2025, March 29). *5 Fakta Bareskrim Bongkar Kasus Scam Kripto Internasional Rp 105 Miliar*. https://news.detik.com/berita/d-7832118/5-fakta-bareskrim-bongkar-kasus-scam-kripto-internasional-rp-105-miliar

Niman, S., Parulian, T. S., & Rothhaar, T. (2023). Online love fraud and the experiences of indonesian women: A qualitative study. *International Journal of Public Health Science (IJPHS)*, *12*(3), 1200. https://doi.org/10.11591/ijphs.v12i3.22617

Novianty, D. (2025, March 10). *Riset: Lebih dari 500 Ribu Serangan Phishing pada Bisnis di Asia Tenggara 2024, Indonesia Nomor Dua di Asia Tenggara*. https://www.suara.com/tekno/2025/03/10/110259/riset-lebih-dari-500-ribu-serangan-phishing-pada-bisnis-di-asia-tenggara-2024-indonesia-nomor-dua-di-asia-tenggara

Otoritas Jasa Keuangan. (2024, August 2). *Joint Press Release: OJK And Statistics Indonesia Present National Survey On Financial Literacy And Inclusion 2024 Findings*. https://ojk.go.id/en/berita-dan-kegiatan/siaran-pers/Pages/OJK-And-Statistics-Indonesia-Present-National-Survey-On-Financial-Literacy-And-Inclusion-2024-Findings.aspx

Paganini, P. (2024, January 12). Vast Voter Data Leaks Cast Shadow Over Indonesia 's 2024 Presidential Election. *Security Affairs*. https://securityaffairs.com/157357/deep-web/hackers-data-leak-indonesia-elections.html

Paramitha, D. D. (2022, Desember | 20.15 WIB). *Cerita Kasus Penipuan Berkedok Bea Cukai, Beli Barang via Medsos Berujung Pemerasan | tempo.co*. Tempo. https://www.tempo.co/ekonomi/cerita-kasus-penipuan-berkedok-bea-cukai-beli-barang-via-medsos-berujung-pemerasan-236913

Peraturan Menteri Komunikasi dan Informatika Nomor 5 Tahun 2020. (n.d.). Retrieved April 17, 2025, from https://jdih.komdigi.go.id/produk_hukum/view/id/759/t/peraturan+menteri+komunikasi+dan+informatika+nomor+5+tahun+2020

Peraturan OJK No. 12 Tahun 2024. (n.d.). Database Peraturan | JDIH BPK. Retrieved April 17, 2025, from http://peraturan.bpk.go.id/Details/301737/peraturan-ojk-no-12-tahun-2024

Post, T. J. (n.d.). *Gambir Police nab 20 suspects in love scam targeting foreigners—Jakarta*. The Jakarta Post. Retrieved April 17, 2025, from https://www.thejakartapost.com/indonesia/2025/01/29/gambir-police-nab-20-suspects-in-love-scam-targeting-foreigners.html

Prabowo, H. Y. (2023). When gullibility becomes us: Exploring the cultural roots of Indonesians' susceptibility to investment fraud. *Journal of Financial Crime*, *31*(1), 14–32. https://doi.org/10.1108/JFC-11-2022-0271

Putra, F. H., Suhardjanto, D., Trinugroho, I., & Arifin, T. (2024). Overcoming Barriers to Inclusion: The Role of Financial Literacy and Digital Divide in Expanding Financial Access in Indonesia. *Journal of Ecohumanism*, *3*(8), Article 8. https://doi.org/10.62754/joe.v3i8.5608

R, M. A. (2016, February 1). *Polda Metro Bekuk Sindikat Penipuan Online Jaringan China di Surabaya*. detiknews. Retrieved April 17, 2025, from https://news.detik.com/berita/d-3132271/polda-metro-bekuk-sindikat-penipuan-online-jaringan-china-di-surabaya

Ratusan mahasiswa IPB jadi korban penipuan, kini diteror penagih pinjol—'Sudah jatuh, tertimpa tangga'. (2022, November 17). BBC News Indonesia. https://www.bbc.com/indonesia/articles/c165dj3lzl2o

Redaksi, T. (2025, March 11). Dominasi Korban Penipuan Online Adalah Perempuan Selama Tahun 2022. *BaliNews.Id*. https://balinews.id/dominasi-korban-penipuan-online-adalah-perempuan-selama-tahun-2022/

Saleh, A. I., & Winata, M. D. (2023). *Indonesia's Cyber Security Strategy: Problems and Challenges*. 1675–1696. https://doi.org/10.2991/978-2-38476-152-4_169

Saleh, G. (2022). JURIDICAL ANALYSIS OF THE CRIME OF ONLINE STORE FRAUD IN INDONESIA. *Jurnal Hukum Dan Peradilan*, *11*(1), Article 1. https://doi.org/10.25216/jhp.11.1.2022.151-175

Sasongko, J. P. (2016, February 22). *Polisi Tangkap Kelompok Penipu Jual-Beli Online*. nasional. Retrieved April 17, 2025, from https://www.cnnindonesia.com/nasional/20160222161552-12-112638/polisi-tangkap-kelompok-penipu-jual-beli-online

Setiawati, S. (2024, September 28). *Uang Kandas, Cinta Pun Melayang: Love Scamming Buat Rugi Rp600 Juta!* CNBC Indonesia. Retrieved April 17, 2025, from https://www.cnbcindonesia.com/research/20240922174024-128-573665/uang-kandas-cinta-pun-melayang-love-scamming-buat-rugi-rp600-juta

Shaw, K. (2023, December 13). Bank Central Asia says "Don't Know? Say No!" in new FCN Creative - Flock campaign. *Campaign Brief Asia*. https://campaignbriefasia.com/2023/12/13/bank-central-asia-says-dont-know-say-no-in-new-fcn-creative-flock-campaign/

Shofa, J. N., & Muslim, A. (2024, October 15). *Indonesia's Internet Users More than Double Over Past Decade*. Jakarta Globe. https://jakartaglobe.id/tech/indonesias-internet-users-more-than-double-over-past-decade

Sistem Penalti Pelanggaran Ketentuan Transaksi. (n.d.). Tokopedia Care. Retrieved April 17, 2025, from https://www.tokopedia.com/help/article/apa-itu-sistem-penalti-pelanggaran-ketentuan-transaksi

Syafaruddin, M. (2024, June 20). *Perempuan Asal Surabaya Jadi Korban Penipuan Online, Rp115 Juta Melayang*. https://www.suarasurabaya.net/kelanakota/2024/perempuan-asal-surabaya-jadi-korban-penipuan-online-rp115-juta-melayang/

Tentang Kami—IASC. (n.d.). Retrieved April 17, 2025, from https://iasc.ojk.go.id/about-us

Tentang Shopee Garansi 100% Ori dan Keuntungannya | Pusat Edukasi Penjual Shopee Indonesia. (n.d.). Retrieved April 17, 2025, from https://seller.shopee.co.id/edu/article/6840

Teresia, A. (2024, July 12). Indonesia says it has begun recovering data after major ransomware attack. *Reuters*. https://www.reuters.com/technology/cybersecurity/indonesia-says-it-has-begun-recovering-data-after-major-ransomware-attack-2024-07-12/

UNODC. (2023). *Casinos, cyber fraud, and trafficking in persons for forced criminality in Southeast Asia*. https://www.unodc.org/roseap/uploads/documents/Publications/2023/TiP_for_FC_Policy_Report.pdf

UU No. 1 Tahun 2024. (n.d.). Database Peraturan | JDIH BPK. Retrieved April 17, 2025, from http://peraturan.bpk.go.id/details/274494/uu-no-1-tahun-2024

Viral Penipuan Modus File APK "Surat Panggilan Polda Metro Jaya." (2024, April 10). https://www.cnnindonesia.com/teknologi/20240410151720-185-1085028/viral-penipuan-modus-file-apk-surat-panggilan-polda-metro-jaya

Watters, P. (2024). *Exposing the Dark Side: Scams and Cybersecurity Risks in Indonesia's Illicit Sports Streaming Scene* (SSRN Scholarly Paper 4954969). Social Science Research Network. https://doi.org/10.2139/ssrn.4954969

Yulia, R., & Sofian, A. (2024). *E-Wallet Misuse in Online Child Prostitution Transactions; How Does Indonesian Law Respond?* 36–43. https://doi.org/10.2991/978-2-38476-325-2_5

Yusriadi, Y., Rusnaedi, R., Siregar, N. A., Megawati, S., & Sakkir, G. (2023). Implementation of artificial intelligence in Indonesia. *International Journal of Data and Network Science*, *7*(1), 283–294. https://doi.org/10.5267/j.ijdns.2022.10.005

Zhu, N. (2018, June 18). *A day in the life of Go-Jek's VP of fraud*. Tech in Asia. https://www.techinasia.com/talk/day-life-vp-fraud-gojek

Online Fraud and Scams in Japan

Daichi Ishii

5

Online Fraud and Scams in Japan

Daichi Ishii[15]

PATTERNS & TRENDS

From 2024 into 2025, both the incidence and the monetary losses of online fraud in Japan have reached record highs. Provisional statistics released by the National Police Agency show that total losses from "special fraud" in 2024 amounted to ¥ 721.5 billion, a year-on-year increase of 59.4 percent, while recognized cases rose to 20,987; there is still no sign of decline as 2025 unfolds (National Police Agency 2025a).

Among these crimes, "SNS-based investment and romance scams," in which perpetrators impersonate celebrities on social media and promise high returns, accounted for 9,265 cases and ¥ 114.1 billion between January and November 2024, making them the chief driver of the overall surge (Itakura 2025). The same scam type has continued at pace in 2025: losses in March alone reached ¥ 5.53 billion, up ¥ 2.37 billion from the previous month and sharply reversing the decline seen through February (National Police Agency 2025a).

According to the Anti-Phishing Council Japan, 2023 set an all-time record with 1.19 million reported phishing incidents; in 2024 the monthly total has repeatedly exceeded 180,000, indicating a persistent upward trend (Anti-Phishing Council 2024).

The contact channels used by scammers have shifted dramatically. In 2023 roughly half of investment scams began via banner advertisements, but after mass takedowns by platforms, scams initiated through direct messages (DMs) overtook ad-based fraud from July 2024 onward. Matching apps, Instagram, and Facebook now rank as the top three channels through which victims are approached (National Police Agency 2024a). Because DMs draw victims into semi-closed spaces that are harder to monitor, this channel shift is cited as a cause of soaring losses.

Real-world cases confirm the escalating damage. In Ibaraki Prefecture, a woman in her seventies who was lured through a LINE group transferred ¥ 809 million in just a few weeks (Furusho & Tokonami 2024). In Hokkaido, a man in his seventies lost ¥ 240 million after being deceived by a DM from someone posing as a "famous analyst" (NHK 2024).

Generative-AI tools are making attacks more sophisticated. A McAfee risk survey found that deep-fake images and names of influential Japanese figures such as Fusaho Izumi (a Japanese politician) and Takafumi Horie (a Japanese famous entrepreneur) are being widely reused, and it concludes that Japanese users' high level of trust in local celebrities amplifies scam persuasiveness (McAfee 2024). Scammers have already been detected playing synthetic voices of famous investors during LINE calls to coax victims into investing, bringing "personal appearances" generated by AI into the real world (Kansai Television 2024).

The victim profile is changing as well. Whereas people aged 65 and over once comprised the vast majority, men in their fifties now account for 29 percent of SNS investment-scam victims, followed by those in their sixties at 27 percent; being defrauded is no longer limited to the elderly (National Police Agency 2024a). Geographically, roughly 65 percent of losses are concentrated in seven prefectures—

[15] Research Consultant, Safer Internet Lab

including Tokyo, Kanagawa, and Osaka—highlighting the vulnerability of major metropolitan areas (National Police Agency 2025b).

Taken together, Japan's online-fraud landscape is defined by four overarching trends:

1. Record-high total and per-case losses
2. A shift toward closed contact channels
3. Greater technical sophistication enabled by generative AI
4. An expanding age range of victims

PATTERNS & TRENDS

Although Japan has rolled out a variety of regulations and administrative measures to address online fraud, the growth in both losses and prosecutions continues to outpace legislation and policy implementation.

In April 2024, the Ministry of Economy, Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC) issued *AI Service Provider Guidelines 1.0*, positioning deep-fake risk management at each stage of development, provision, and use (Ministry of Economy, Trade and Industry and Ministry of Internal Affairs and Communications 2024). An AI Bill subsequently cleared the House of Representatives on 24 April 2025, establishing a framework that allows the government to request investigative cooperation from service providers when serious incidents occur and—via an accompanying resolution—assigns explicit state responsibility for countering deep-fake pornography (Murai 2025).

The government also adopted a *Comprehensive Strategy to Protect Citizens from Fraud* in June 2024, designating SNS-based investment scams as a top-priority offense and allocating an extra ¥ 650 million in the supplementary budget (Prime Minister's Office 2024). According to National Police Agency statistics, 232 SNS investment-and-romance scams were prosecuted and roughly 1,000 related bank accounts were frozen during 2024 (National Police Agency 2025b). The Financial Services Agency (FSA) has opened a *Fraudulent Investment Consultation Dial*, although detailed call statistics have yet to be released.

As an anti-phishing measure against spoofed government websites, the Digital Agency is designing a new authenticity icon for public-sector sites, with pilot deployment targeted in fiscal-year 2025 (Digital Agency 2025a).

Some local governments are taking their own initiatives. Tottori Prefecture enacted an ordinance effective 1 April 2025 that bans the creation, distribution, or provision of AI-generated pornographic images using real children's faces; violations carry penalties of up to one year's imprisonment or a fine of up to ¥ 500,000 (Tottori Prefecture 2025).

International collaboration is also advancing. At the minister-level UK–Japan Digital Partnership meeting in January 2025, the two countries agreed to co-develop deep-fake verification benchmarks between their respective AI Safety Institutes (Digital Agency 2025b). Japan likewise joined the United States, Australia, and six other countries in co-signing an international advisory that explicitly named the China-linked hacker group APT40, strengthening information-sharing on cryptocurrency-based money-laundering techniques (National Police Agency 2024b).

Even so, these efforts alone cannot fully address the rapid evolution of generative-AI-enabled fraud and the anonymity of cross-border transfers via crypto-assets. Agile regulation that keeps pace with technological advances, together with real-time mechanisms for tracking damage, remains essential.

PRIVATE-SECTOR INITIATIVES

Major platforms are likewise strengthening their defenses against online fraud. In July 2024, Meta further disclosed a pilot program for the Japanese market that withholds ad revenue for 90 days and refunds it to victims; during that pilot the company disabled the same **5.27 million** scam ads and **5,400** accounts between 5 March and 1 June 2024 (Meta 2024). **LINE Yahoo** has launched a cross-functional project that flags unverified accounts and tightens ad screening, outlining the measures on its corporate blog (LY Corporation 2025).

The trend is not confined to Japan. Google's **"Ads Safety Report 2024"** states that the company permanently terminated **more than 700,000** malicious advertiser accounts worldwide and cut reports of celebrity deep-fake ads by **90 percent** year on year (Google 2024).

On the financial-infrastructure side, the **Japanese Bankers Association** put a **fund-transfer-freeze API** into operation in October 2024, enabling real-time blocks through bank–police cooperation. In the crypto-asset arena, the industry body **JVCEA** is preparing to introduce **real-time KYT (Know Your Transaction)**, with joint investigation of suspicious transfers emerging as a key challenge (Chainalysis 2025).

Prominent individuals are also pushing back: billionaire **Yusaku Maezawa** (a Japanese famous entrepreneur) and others have announced plans to sue Meta, marking new cases in which celebrities exploited in scams take legal action (Toyo Keizai Online 2024).

Yet, as noted earlier, fraud now originates not only from banner ads but increasingly from **direct messages**. Capturing the full picture of these harder-to-see schemes will be a critical challenge going forward.

RECOMMENDATIONS

To shift Japan's response to online fraud from *after-the-fact crackdowns* to *anticipatory prevention*, three pillars should be advanced in parallel.

Pillar 1 – Stronger platform duties and built-in advanced detection

Platforms could be required to auto-forward metadata on suspected scam DMs to the National Police Agency's special task force. In cooperation with the entertainers' guild, a **"celebrity-image whitelist"**—authentic photos hashed to boost AI-based verification—could further raise detection accuracy. On the user side, the moment an ad or DM triggers a fraud signal, platforms could display "typical scam phrases" and real-time loss statistics, delivering behavioral-science-based warnings that stop victims before they act.

Pillar 2 – Open, API-linked data infrastructure

An infrastructure that continuously links public- and private-sector data via APIs and shows it in a public dashboard is worth considering. For example, the Digital Agency might host a weekly-updated dashboard of KPIs—loss amounts, case counts, ad takedowns—pulled from the National Police Agency, Financial Services Agency, and major platforms, while supplying reusable APIs for researchers and journalists.

Pillar 3 – Education

Although the payoff is long-term, bolstering educational initiatives is essential for raising society-wide resilience. Programs that let students *experience* AI-powered scam scenarios through hands-on learning could be one promising approach.

Interlocking these multilayered measures would position Japan to build one of the world's strongest defense nets against "AI-era" fraud. A concerted effort by government, industry, and academia is now called for, ensuring that the next generation can enjoy digital life in safety.

REFERENCES

Anti-Phishing Council Japan. 2024. 「フィッシングレポート 2024」. https://www.antiphishing.jp/report/phishing_report_2024.pdf

Chainalysis. 2024. 「日本における暗号資産のマネーロンダリング: 日本の視点から見たグローバルの共通問題」. https://www.chainalysis.com/blog/crypto-money-laundering-japan-japanese/.

Chainalysis. 2025. 「2024 年の暗号資産詐欺: 詐欺産業が AI を活用し巧妙化する中、ロマンス詐欺は前年比でほぼ 40 %増加」. https://www.chainalysis.com/blog/2024-pig-butchering-scam-revenue-grows-yoy-japanese/.

Digital Agency. 2025a. 「公的サイトの真正性確認に関する新アイコン導入方針」. https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/5ecac8cc-50f1-4168-b989-2bcaabffe870/8b045435/20250315_auth_icon.pdf.

Digital Agency. 2025b. 「第3回日英デジタルパートナーシップ政務級会合の結果」. https://www.digital.go.jp/news/02d07f5c-0301-48ba-9150-7119457195f8.

Financial Services Agency. 2024. 「『詐欺的な投資に関する相談ダイヤル』の開設について」. https://www.fsa.go.jp/news/r5/sonota/20240619/toshisagi.html.

Financial Services Agency. 2025. 「暗号資産に関連する制度のあり方等の検証」. https://www.fsa.go.jp/news/r6/sonota/20250410_2/crypto_dp.html.

Furusho, Noboru, and Koichi Tokonami. 2025. 「SNS型詐欺など昨年の被害額は年比 2 倍超に　被害は若者にも拡大」. *Asahi Shimbun Digital*. https://www.asahi.com/articles/AST1P45P7T1PUJHB001M.html.

Google. 2025. “2024 Ads Safety Report.” https://services.google.com/fh/files/misc/ads_safety_report_2024.pdf.

Itakura, Daichi. 2025. 「特殊詐欺と SNS 型投資・ロマンス詐欺　被害は 2 千億円、過去最悪」. *Asahi Shimbun Digital*. https://www.asahi.com/articles/AST253Q7TT25UTIL01NM.html.

Kansai Television. 2024. 「AIを使ったフェイク動画・音声で　堀江貴文氏かたる投資勧誘　SNS投資被害で5260 万円被害」. https://www.ktv.jp/news/feature/240415/.

LY Corporation. 2025. 「『LINE』における詐欺行為の撲滅に向けた取り組み」. https://www.lycorp.co.jp/ja/story/20250331/snsscam.html. https://www.lycorp.co.jp/ja/story/20250404/snsscam2.html

McAfee. 2024. 「マカフィー、『2024 年オンライン詐欺で悪用されやすい日本の著名人 TOP10』を発表」. https://www.mcafee.com/ja-jp/consumer-corporate/newsroom/press-releases/2024/20241024.html.

Meta. 2024. 「詐欺広告に対する取り組み強化について」. https://about.fb.com/ja/news/2024/07/updates_on_tackling_scams/.

Ministry of Economy, Trade and Industry. 2025. 「2024 年度 デジタルプラットフォームの透明性に関する評価報告書」. https://www.meti.go.jp/policy/mono_info_service/digitalplatform/evaluation.html.

Ministry of Economy, Trade and Industry and Ministry of Internal Affairs and Communications. 2024. 「AI事業者ガイドライン（第 1.0 版）」. https://www.meti.go.jp/press/2024/04/20240419004/20240419004.html.

Murai, Naoko. 2025. 「AI法案が衆院可決　付帯決議でディープフェイクポルノ対策求める」. *Asahi Shimbun Digital*. https://www.asahi.com/articles/AST4S1GLCT4SULFA009M.html.

National Consumer Affairs Center of Japan. 2024. 「SNS をきっかけとして…勧誘される金融商品・サービスの消費者トラブルが急増」. https://www.kokusen.go.jp/pdf/n-20240529_1.pdf.

National Police Agency. 2024a. 「令和 6 年 11 月末における SNS 型投資・ロマンス詐欺の認知・検挙状況等について」. https://www.npa.go.jp/bureau/criminal/souni/sns-romance/sns-touroma2024.pdf.

National Police Agency. 2024b. 「豪州主導の APT40 グループに関する国際アドバイザリーへの共同署名について」. https://www.npa.go.jp/bureau/cyber/koho/caution/caution20240709.html.

National Police Agency. 2025a. 「令和 7 年 3 月末における特殊詐欺及び SNS 型投資・ロマンス詐欺の認知・検挙状況等について」. https://www.npa.go.jp/bureau/safetylife/sos47/new-topics/250428/02.html.

National Police Agency. 2025b. 「令和６年における特殊詐欺及びＳＮＳ型投資・ロマンス詐欺の認知・検挙状況等について（暫定値版）」. https://www.npa.go.jp/bureau/criminal/souni/tokusyusagi/hurikomesagi_toukei2024.pdf.

NHK (Japan Broadcasting Corporation). 2024. 「SNS投資詐欺で 2 億 4 000 万円余の被害 道内被害で最高額」. https://www3.nhk.or.jp/sapporo-news/20241204/7000071781.html.

Nozomi Sogo Law Office. 2023. 「改正電気通信事業法による外部送信規律（日本版 Cookie 規制）」. https://www.nozomisogo.gr.jp/newsletter/9660.

Prime Minister’s Office. 2024. 「国民を詐欺から守るための総合対策 2.0」. https://www.kantei.go.jp/jp/singi/hanzai/kettei/250422/honbun-1.pdf.

Shinya, Chifumi. 2024. 「『株投資の勉強会』と誘導され 2 億円超の詐欺　利益 18 億円が一転」. *Asahi Shimbun Digital*. https://www.asahi.com/articles/ASSD42S8XSD4IIPE00HM.html.

Tottori Prefecture. 2025. 「知事定例記者会見（2025 年 1 月 14 日）」. https://www.pref.tottori.lg.jp/320891.htm.

Toyo Keizai Online. 2024. 「Facebook で『詐欺広告』が放置され続ける真因」. https://toyokeizai.net/articles/-/750379.

Online Fraud and Scams in the Philippines

Jose Carlos Alexis Bairan,
Queen Cel Oren

6

Online Fraud and Scams in the Philippines

Jose Carlos Alexis Bairan[16], Queen Cel Oren[17]

INTRODUCTION

As with most countries globally, the Philippines experienced an accelerated shift towards digitalization in the past years due to the advancements in technology and the COVID-19 pandemic, transforming many aspects of how Filipinos lived and worked. The digital shift provided an opportunity for people to work, study, and do most activities remotely, while it enabled organizations to continue operating amidst the mobility challenges posed by the COVID-19 pandemic. These two factors encouraged the growth and widespread adoption of various online tools such as digital wallets (e-wallets), digital banking, online payment systems, online shopping, and remote meeting platforms.

However, the promised convenience, connectivity, efficiency, and transparency of digitalization came with increased susceptibility of Filipinos to fraudulent schemes over the internet. In 2024 alone, it was reported by the Central Bank of the Philippines or *Bangko Sentral ng Pilipinas* (BSP) (2024) that around 59.4 percent of internet users have encountered or have fallen victim to online fraud. Perpetrators trick their victims through various methods, including identity theft, fake websites and profiles, digitally-tampered documents, and spammed text messages, among many others. These types of scams are also quickly becoming more problematic with the rise of artificial intelligence, which enabled online scammers to craft and execute more sophisticated scams.

A study on internet use by Meltwater and We Are Social (2024) reported that 91.3 percent of Filipino internet users aged 16 years old and higher use banking, investment, or insurance websites or applications every month. Accordingly, data from the BSP reveal that the number of e-money accounts in the Philippines significantly expanded from 257.5 million in 2022 to 393.6 million accounts in 2023. The value of digital transactions in the country totaled to about PHP 110.5 billion, making up 55.3 percent of the total retail transactions value in 2023 (Bangko Sentral ng Pilipinas, 2024). The sheer volume of digital payments, internet users, e-money accounts, and internet users in the country makes it a very attractive target of artificial intelligence (AI)-powered online scams.

The growing number and sophistication of scams amidst the pervasive use of digital tools, especially those that involve personal information and financial transactions, underscore that cybersecurity has become an imperative for individuals and organizations. Against this background, this paper aims to provide an overview of the patterns and trends in online scams in the Philippines, the country's regulatory and policy environment, current initiatives and challenges, and recommendations to enhance the country's defenses against various kinds of online scams.

PATTERNS AND TRENDS IN ONLINE SCAMS IN PHILIPPINES

Online scams in the country continue to evolve, with scammers leveraging AI for phishing, voice cloning, ransomware, and deep-fake videos to deceive individuals into revealing personal information or making financial transactions. AI-generated content is also used to craft highly personalized messages targeting

[16] Research Associate, Ateneo Center for Research and Innovation
[17] Research Specialist, Philippine Institute for Development Studies

specific individuals or demographics, making fraudulent communications more difficult to detect. However, identifying whether a scam is AI-enabled remains a challenge.

Online scams occur frequently, with some being cross-border. Some private industries measure both financial and non-financial impacts but usually keep such information confidential. AI-enabled tools are used for scam detection, though specific details are not shared publicly. A significant number of scams take place on social media platforms, and organizations emphasize data privacy and customer security. While public awareness of online scams is high, trust in government protective measures remains relatively low.

Phishing occurs across various platforms, including calls, SMS, emails, phishing links, fraudulent social media pages, e-commerce apps, and lost or stolen phones. Incidents of hacking emails and social media accounts are also common. Pyramid schemes use deceptive tactics, such as incorporating tasks like CAPTCHA encoding, to appear legitimate while relying on recruitment rather than sustainable product sales.

Scammers also exploit social media and messaging apps for romance scams, where they build trust with victims before defrauding them. Package scams deceive individuals into paying fees for non-existent deliveries, often using platforms like Facebook Messenger. As scams continue to adapt and spread, heightened vigilance, stronger cybersecurity measures, and proactive fraud detection remain essential in combating these threats.

SOCIO-ECONOMIC IMPLICATIONS OF SCAMS

Online scams result in substantial financial losses for individuals, businesses, and financial institutions, disrupting economic stability and personal financial security. Fraudsters continuously exploit weaknesses in digital systems, employing deceptive tactics such as phishing, business email compromise, hacking, and romance scams to manipulate victims into divulging sensitive information or transferring funds. These scams evolve with technological advancements, making them increasingly difficult to detect and prevent.

Certain groups are particularly vulnerable to online scams, with senior citizens and low-income individuals facing a heightened risk due to limited digital literacy and a lack of awareness about cybersecurity threats. Many elderly victims fall prey to scams disguised as urgent financial requests or fraudulent investment opportunities. At the same time, low-income individuals may be enticed by fake job offers or loan scams promising quick financial relief. These deceptive practices not only deplete personal savings but also contribute to emotional distress and economic hardship, further widening social inequalities.

Businesses, too, are severely impacted by online scams, experiencing both direct financial losses and secondary consequences that affect long-term stability. Cybercriminals often target companies through sophisticated social engineering schemes, compromising business operations and eroding customer trust. Beyond immediate financial setbacks, organizations suffer reputational damage that may result in lost clientele, decreased investor confidence, and increased regulatory scrutiny. Resources allocated to mitigating cyber fraud—such as enhanced security measures, legal fees, and fraud investigations—divert funds away from business expansion, innovation, and job creation, stifling overall economic progress.

The repercussions of online scams extend far beyond financial losses, influencing public confidence in digital financial services. Widespread cybersecurity concerns deter individuals from fully participating in digital transactions, limiting the adoption of online banking, digital payments, and other fintech solutions. Victims who experience financial fraud often reduce discretionary spending due to diminished trust in

online platforms, leading to lower consumer engagement and weaker investment activity. This hesitancy directly affects government-led financial inclusion initiatives aimed at expanding access to credit, insurance, and investment opportunities for Filipinos. A lack of trust in digital financial systems hampers the country's transition toward a more digitally connected economy, slowing economic growth and innovation.

Compounding this issue is the increasing misuse of AI in executing scams, allowing fraudsters to refine their tactics with greater sophistication. AI-driven phishing attacks, deepfake scams, and automated fraud schemes make it more difficult to differentiate between legitimate and fraudulent transactions as scammers continue to leverage AI to enhance deception, traditional detection methods become less effective, necessitating continuous advancements in cybersecurity defenses.

The persistent rise in online fraud highlights the need for ongoing research into emerging scam trends and the effectiveness of awareness initiatives. Understanding how scams evolve, assessing the effectiveness of current preventive measures, and enhancing public preparedness are crucial in mitigating financial crimes. Without proactive intervention, the economic and social consequences of cyber fraud will continue to escalate, reinforcing skepticism toward digital financial services and hindering the country's progress in building a secure and inclusive digital economy.

POLICY ASSESSMENT IN ADDRESSING ONLINE FRAUD AND SCAMS

The Role of Key Stakeholders in Addressing Scams in Philippines

There are several Philippine government offices and law enforcement agencies with various functions and responsibilities that are relevant to ensuring that Filipinos are safe and secure from online scams. The frontline of the country's financial system, the BSP, formulates, implements and monitors compliance to policies and regulations that govern BSFIs, which are the prime targets of online scammers.

The Department of Information and Communications Technology (DICT) is the primary government entity involved in policy making, planning, coordinating, implementing, and administering the development of the national information and communications technology (ICT) sector. The DICT leads the crafting and implementation of the NCSP to set standards that aim to enhance cybersecurity across government agencies and critical infrastructures. It also oversees the National Computer Emergency Response Team (NCERT), which handles cybersecurity incident responses. Other offices attached to the DICT with cybersecurity-related mandates are as follows:

- **Cybercrime Investigation and Coordinating Center (CICC)** - responsible for developing and employing coordinating mechanisms that allow law enforcement agencies, telecommunications industry, and other key stakeholders to work together in cybercrime prevention, investigation, and enforcement.
- **National Privacy Commission (NPC)** - Ensures the compliance of individuals and organizations to the Data Privacy Act and monitors possible data privacy violations in the collection, storage, processing, and use of personal data, which is usually linked to online scams.
- **National Telecommunications Services (NTC)** - regulates all telecommunications services and provides reporting mechanisms for telecommunications service providers and users to report incidents related to online scams.

There are also several law enforcement agencies involved in ensuring that online scam incidents in the country are appropriately addressed. The Office of Cybercrime under the Department of Justice (DOJ) acts as the focal government office for investigating and prosecuting cybercrime cases, including online

scams, under the Cybercrime Prevention Act. Meanwhile, the Philippine National Police's Anti-Cybercrime Group enforces laws against cybercrimes, receives reports on online scams and enforces entrapment operations, while the National Bureau of Investigation's Cybercrime Division conducts investigations, gathers evidence, and runs forensic analyses on cybercrime cases.

To keep up with the expansion of the digital economy, particularly e-commerce, the E-Commerce Bureau under the Department of Trade and Industry (DTI) was also established through the passage of the Internet Transactions Act in 2023. With the Bureau's mandate to formulate and oversee the policies for e-commerce, the Philippine government is expected to better regulate online transactions, enhance consumer protection, and foster a high-trust environment between consumers and businesses.

Table 6.1 Summary Matrix of Philippine Government Offices with Cybersecurity-Related Roles

Authority	Role in Online Scams and Fraud
Policy Formulation and Oversight	
Department of Information and Communications Technology (DICT) *Includes the following offices:* 1. Cybercrime Investigation and Coordinating Center (CICC) 2. National Telecommunications Commission (NTC) 3. National Privacy Commission (NPC)	• Oversees policies and programs related to the development of the national ICT sector, data privacy, security, and confidentiality • Formulates cybersecurity policies that aim to prevent, address, and minimize cyber threats and attacks • Provides countermeasures to address domestic and transnational cyber incidents • Monitors cybercrime cases handled by law enforcement agencies
Bangko Sentral ng Pilipinas (BSP)	• Formulates, implements, and monitors policies and regulations to guide BSFIs in cybersecurity-related concerns • Assists in mediating complaints between BSFIs and consumers
Department of Trade and Industry (DTI)	• Formulates policies and monitors and oversees e-commerce transactions • Handles consumer complaints about unfair trade practices
Department of Justice – Office of Cybercrime (DOJ-OOC)	• Handles the prosecution of cybercrime cases, including online scams, that violate the provisions of the Cybercrime Prevention Act • Responsible for international cooperation on legal assistance and extradition, which may involve resolving issues related to cross-border fraudulent transactions
Anti-Money Laundering Council (AMLC)	• Although it has no mandate to prevent online scams, it provides initiatives to create infographics for public awareness.
Law Enforcement Agencies	
Philippine National Police – Anti-Cybercrime Group	• Enforces laws, conducts cybercrime investigations, including online scams, and raises public awareness against online fraud
National Bureau of Investigation - Cybercrime Division	• Investigates investment scams, cybercrime, and other types of online scams

Significant Policy Developments

To combat the growing volume and sophistication of online scams, the Philippine government has several laws, regulations, and plans in place that build a strong cybersecurity system that aim to protect businesses and consumers from these crimes. Over the past decade, the Philippines has made significant progress in tightening measures not only to enhance the resilience of digital transactions against online scams, but also to protect consumers' data.

For instance, in 2012, the Philippines was ahead of other Southeast Asian countries in introducing legislation on data privacy and cybercrime. The Cybercrime Prevention Act (RA 10175) criminalizes fraudulent activities facilitated online such as hacking, identity theft, and online fraud, including scams that utilize AI technologies. On the other hand, the Data Privacy Act (RA 10173) regulates the storage, processing, and use of personal information. It mandates organizations to comply with the set security measures to protect personal sensitive data from unauthorized access, disclosure, or loss to reduce the risk of data breaches that can be used maliciously for scams and other cybercrimes.

Anti-Money Laundering Act (AMLA) (RA 9160) or the also serves as key legislative tools in combating online scams, AI-driven fraud, and cross-border cybercrimes. The AMLA establishes the legal framework for monitoring financial transactions, detecting suspicious activities, and working with international organizations to address financial crimes linked to scams. These legal frameworks provide a foundation for a coordinated response to evolving cyber threats.

More recently, new legislation on consumer protection and cybersecurity such as the "Financial Products and Services Consumer Protection Act" (FCPA) (RA 11765) and the "Anti-Financial Account Scamming Act (AFASA)" (RA 12010) to fortify the consumers and financial industry's line of defense against online scams. FCPA empowered the BSP, along with other financial regulators, to implement rules and mechanisms to address consumer complaints. It also encourages financial institutions to establish accessible and reporting mechanisms for consumers to report anomalies. Meanwhile, the AFASA compels all BSP-supervised financial institutions (BSFIs)[18] to adopt more rigorous measures to protect consumers. It reinforces the responsibility of BSFIs to employ proper fraud management systems, infrastructure and security monitoring, multi-factor authentication, and user enrollment and verification processes. The legislation also sought to enhance coordination between financial institutions and law enforcement agencies by giving the BSP power to investigate suspicious transactions and share the results with relevant law enforcement agencies.

These laws are complemented by the recent launch of the Financial Services Cyber Resilience Plan (FSCRP), which will serve as the country's roadmap and framework to strengthen the financial system's defenses against cyberthreats and cybercrimes in the next 5 years. The FSCRP highlights the need for stronger information sharing and collaboration between BSFIs, government agencies, and other stakeholders through cyber threat and incident reporting, inventory, and mapping.

Moreover, the BSP issued key circulars to reinforce financial security. It issued BSP Circular No. 808, s. 2013, establishing IT risk management guidelines, requiring financial institutions to implement stringent security controls; BSP Circular No. 982, s. 2017, strengthening information security to prevent fraud; and BSP Circular No. 1019, s. 2018, mandating all BSFIs to report cyber-related incidents.

Other relevant policies and regulations help combat online scams by strengthening cybersecurity and consumer protection. Internet Transactions Act of 2023 (RA 11967) aims to enhance protection in online transactions and improve consumer trust by regulating businesses and consumers engaged in online

[18] Includes banks, quasi-banks, pawnshops, foreign exchange dealers, money changers, remittance agents, electronic money issuers or e-wallets, and non-stock savings and loan associations.

transactions. This involves creating an e-commerce bureau in charge of maintaining a database of online businesses, ensuring dispute resolution mechanisms, and creating a code of conduct for both businesses and consumers. The adoption of the National Cybersecurity Plan 2023-2028 aligns with the Philippine Development Plan (PDP) to enhance cybersecurity, build a more skilled workforce, and strengthen policy frameworks for a safer digital environment. The bill on the Cybersecurity Act, which seeks to enable the government to keep up with major advancements in technology and cybersecurity such as AI, critical information infrastructures, and digital assets, is still pending in the 19th Congress.

Challenges to Combat Online Scams

Cybercriminals continue to develop more sophisticated methods, making it essential for businesses and other entities handling financial transactions to remain vigilant in protecting both themselves and their clients. Some government agencies provide infographics on their websites to inform the public about current and emerging threats.

Campaigns to increase public awareness play a key role in helping individuals recognize and report scams. The effectiveness of scam awareness efforts can be gauged through surveys and feedback, which help determine people's understanding of scam tactics, their ability to recognize red flags, and their reactions to potentially fraudulent situations. As digital financial transactions grow, improving data protection measures, enhancing information-sharing systems, and implementing timely interventions are necessary to strengthen anti-fraud efforts.

Institutions are adopting various IT solutions, including AI-driven tools to prevent, detect, and mitigate cyber threats. Regular cybersecurity advisories and awareness campaigns are disseminated among staff to minimize risks and safeguard IT infrastructure. Additionally, security awareness infographics are shared to educate personnel about emerging scams. Advanced perimeter cybersecurity tools are also deployed to monitor and counter malicious emails and traffic, addressing both conventional and AI-enabled threats.

Evaluating the effectiveness of these anti-scam initiatives requires assessing awareness programs, tracking scam prevalence, and analyzing regulatory measures. AI-driven fraud detection is gaining traction, enhancing transaction monitoring by identifying suspicious patterns and anomalies in real-time. While AI adoption remains limited, its potential to improve detection accuracy, incident response times, and regulatory compliance underscores the need for stronger implementation strategies. However, the lack of AI-related governance policies and skilled professionals, particularly in regional agencies, remains a challenge. Strengthening enforcement and investing in AI-driven security measures are essential to mitigating the growing risks of digital fraud in the Philippines.

A significant challenge in combating online scams is the validation of social media accounts and the spread of misinformation. The Philippines has limited mechanisms to hold major platforms such as Meta, X, and TikTok accountable for the proliferation of fraudulent accounts and false information. Additionally, there are resource limitations in providing public information to combat online scams and other fraudulent activities.

Cross-border online scams remain a growing concern despite ongoing international cooperation, including the use of mutual legal assistance treaties (MLATs). However, the increasing sophistication and global nature of these scams require more streamlined and efficient collaboration between authorities. Strengthening real-time information-sharing mechanisms is necessary to address the rapidly evolving tactics used by cybercriminals.

CONCLUSION AND RECOMMENDATIONS

In the face of more sophisticated cyber threats, particularly the escalating prevalence of emerging fraudulent scams, the Philippines must adopt various approaches to enhance its cybersecurity defenses. Elevating public awareness is of paramount importance. Comprehensive public awareness campaigns are essential to instill a culture of vigilance against online scams. Although detecting online fraud through observation alone is difficult, timely reporting and informing potential victims *through social media platforms* can prevent further harm. These campaigns should go beyond mere warnings, providing practical, actionable tips for identity protection and scam detection, empowering citizens to safeguard their personal information. Fostering collaborative action is also crucial. A robust cybersecurity ecosystem necessitates a strong partnership between government agencies, private sector businesses, and civil society organizations. This collaborative approach facilitates the seamless sharing of threat intelligence, the development of effective countermeasures, and the encouragement of timely reporting of cybercrimes and suspected fraudulent activities. This synergy ensures a unified front against cyber threats.

Empowering Small and Medium-sized Enterprises (SMEs) is vital, as they are often the most vulnerable to cyberattacks. Targeted support and resources must be provided to enhance their cybersecurity capabilities. This includes guidance on prevention strategies, mitigation tactics utilizing artificial intelligence, and access to affordable, yet effective, cybersecurity solutions. The rapid advancement of AI necessitates the urgent development of a comprehensive AI governance framework. This framework must strike a delicate balance between fostering innovation and ensuring responsible deployment. It should address critical ethical considerations, safeguard data privacy, and mitigate potential security risks, ensuring that AI benefits society without compromising its safety. Strengthening regulatory enforcement is crucial. Existing cybersecurity laws, such as the Cybercrime Prevention Act of 2012 (RA 10175) and the Data Privacy Act of 2012 (RA 10173), must be rigorously enforced. This requires a substantial investment in cutting-edge technology, recruiting and training highly skilled personnel, and streamlining enforcement processes.

Empowering the Local Government Units (LGUs) is also critical for a universally secure cyberspace. Cybersecurity efforts must extend beyond Metro Manila, reaching all corners of the country. Investing in LGUs' cybersecurity capabilities ensures effective response and prevention efforts are implemented nationwide. A significant increase in resource allocation is necessary. The government must allocate sufficient funds to acquire state-of-the-art technological tools and to develop a highly skilled workforce capable of effectively combating cybercrime. This investment is essential to building a resilient and secure digital environment for the Philippines. Furthermore, it is essential to encourage further studies that continuously monitor and evaluate emerging trends in AI-driven scams to enhance prevention and mitigation strategies.

REFERENCES

Bangko Sentral ng Pilipinas. (2024). *Financial services resilience plan 2024-2029: Fortifying cyber frontier for BSP-supervised financial institutions.* Bangko Sentral ng Pilipinas, Technology Risk and Innovation Supervision Department.

Bangko Sentral ng Pilipinas. (2024). *Traversing new heights: The future is digital: 2023 status of digital payment in the Philippines.* Bangko Sentral ng Pilipinas.

Financial Inclusion Steering Committee. (2023). *2023 Annual report: National strategy for financial inclusion.* Bangko Sentral ng Pilipinas.

Meltwater & We Are Social. (2024). *Digital 2025 Global overview report: The essential guide to the world's connected behaviours.* Meltwater & We Are Social. From http://www.meltwater.com/en/global-digital-trends

Online Fraud and Scams in Singapore

Joanna Octavia

7

Online Fraud and Scams in Singapore

Joanna Octavia[19]

INTRODUCTION

In recent years, Singapore has seen a surge of online scams, with cybercriminals using increasingly sophisticated tactics to defraud individuals and businesses. As a high-income market, Singapore is attractive for scammers, who target the affluent population and extensive digital connectivity. In 2024, Singapore recorded over 50,000 scam cases, equating to approximately one in every 100 people in Singapore falling victim (Abraham et al., 2024). According to the Global Anti-Scam Alliance (GASA), victims in Singapore suffered the highest average financial losses globally, surpassing victims in Switzerland and Austria (Wong, 2024b). The increasingly complex nature of these scams, combined with advancements in technology and the widespread use of AI, has exacerbated the problem.

As one of the most globally connected markets, Singapore's high levels of connectivity, along with its well-developed financial infrastructure and international banking links, make it an attractive target for scammers. In 2024, Singapore retained its position as the world's most financially inclusive market for the third consecutive year, reflecting the country's robust financial systems and widespread access to financial services (Principal, 2024). The country's strong digital connectivity further supports this status, with 99 percent of households connected to the internet and 96 percent have smartphones in 2024 (IMDA, n.d.). However, the country's push toward a cashless society, supported by the widespread use of digital wallets, QR code payments and contactless transactions, has introduced new vulnerabilities that scammers can capitalise on. Singapore's digital economy, which is built on high levels of trust, further meant that people may be less sceptical about schemes that are fraudulent.

Despite having high digital literacy, Singapore remains vulnerable to online scam attacks. According to the Singapore Police Force, self-effected transfers made up 82.4 percent of all reported scam cases in Singapore in 2024, underscoring the significant role of social engineering in these scams (SPF, 2024a). This phenomenon can be attributed to several factors, such as the use of increasingly sophisticated scam tactics, overconfidence in detecting scams, and the exploitation of social values and trusted communication channels by scammers. To combat these challenges, the government is ramping up public-private collaboration between regulators, financial institutions and technology companies. Of note is the increased enforcement powers through the recent passage of the Protection from Scams Bill, continued collaboration between financial institutions and law enforcement, and public awareness raising.

CURRENT TRENDS OF ONLINE FRAUDS AND SCAMS IN SINGAPORE

Social Media and Instant Messaging

Social media platforms and instant messaging apps, both of which are central to communication and commerce, are ranked in the top two contact methods used by scammers (SPF, 2024a). The widespread use of these digital platforms, coupled with Singapore's high rate of social media adoption, made them a fertile ground for scammers to reach a large audience with ease. Three products from Meta, namely Facebook, WhatsApp and Instagram, remain consistently overrepresented among the platforms exploited by scammers (SPF, 2024a).

[19] Researcher, Safer Internet Lab, CSIS Indonesia; Associate Lecturer, University College London

Social media platforms serve as a key tool for scammers to target and deceive potential victims. About 88 percent of Singapore's population use social media, making it a widely used and trusted communication channel (Technode Global, 2024). Among other uses, these platforms have been misused to advertise fake investment schemes; host fake accounts or pages that mimic legitimate organisations to run phishing schemes; and build rapport and relationships with potential victims through social engineering tactics (Koh, 2023; Chia, 2024b; Chiu, 2024; Sim, 2024).

Meanwhile, instant messaging apps are typically used as the more direct and private channel to manipulate victims. They are the most common means for scammers to contact potential victims, with WhatsApp and Telegram as two of the most widely exploited apps (SPF, 2024a). A survey published by GASA (see Abraham et al., 2024) found that nearly three-quarters of respondents were contacted by scammers via WhatsApp, highlighting how scammers exploit the speed, reach, and personal nature of these channels to target potential victims (Abraham et al., 2024). Documented uses of instant messaging apps in online scams in Singapore range from chat rooms perpetrating fraudulent investment schemes, to government official impersonation scams (Tan, 2025; Yasmine, 2025). However, scam-related content on these apps is often beyond direct government oversight due to their encryption and privacy features. Beyond private messaging, these apps also allow for rapid sharing, which can amplify the reach of scam content and make it difficult to contain.

Government Official Impersonation Scams

Government official impersonation scams are one of two types of scams that cause higher losses than others, the other being investment scams (Yasmine, 2025). From January to October 2024, there have been at least 1100 cases reported with total losses amounting to at least SG$ 120 million, almost double to that in the same period in 2023 (MAS, 2025). These scams are highly effective in Singapore primarily due to the strong trust in government institutions and officials, as well as strict law enforcement and law penalties, which made victims fear getting into legal trouble.

Government official impersonation scams are complex schemes that typically involve multiple stages and several scammers. A scammer, posing as a bank officer, would call a potential victim and falsely claim that a victim's credit card was issued, or suspicious transactions had occurred. If the victim denies involvement, the scammer escalates the situation by transferring the call to an accomplice impersonating a Monetary Authority of Singapore (MAS) official or law enforcement officer. Using video calls, the scammers would attempt to build their credibility using fake credentials, agency logos, or fabricated warrant cards and official documents, before shifting the conversation to WhatsApp (MAS, 2024c). Once they are communicating through private messaging channels, the scammers would then pressure victims to transfer money to "safety accounts" under the guise of aiding investigations, ultimately stealing their funds (MAS, 2024c). The complexity of the schemes illustrates the sophisticated tactics employed by scammers. By using various methods and layers of deception, scammers make it difficult for victims to detect that they have been manipulated until it is too late.

Cryptocurrency and Investment Scams

The rising popularity of cryptocurrencies in Singapore provided a way for scammers to lure victims in Singapore into seemingly lucrative cryptocurrency investments, which are fake. The general public's limited understanding of crypto's risks, combined with the hype around digital assets, made them an effective attraction for victims who may not fully grasp the risks involved. Investment scams comprise the highest total amount lost compared to other types of scams in Singapore, reaching at least SG$320.7 million in 2024, underscoring their severity and financial impact (SPF, 2024a).

Investment scams in Singapore leverage the growing interest in cryptocurrencies and the opaque, complex nature of digital assets. After meeting on digital platforms like Facebook, Instagram, Telegram or dating apps, scammers would encourage investment scam victims to invest a small amount of money at the start, which - following a small 'profit' - would be followed by escalation of investment (Police warns of investment scams, 2025). They are asked to open accounts at crypto exchanges and transfer money to the account to buy cryptocurrencies. Once larger amounts of monies or cryptocurrencies have been transferred by the victims to a fraudulent trading platform or the scammers' own wallets, they would begin to experience difficulties in withdrawing their 'investments' (SPF, 2024a; Police warns of investment scams, 2025). In other instances, victims are defrauded by fake advertisements on social media platforms such as Facebook or Instagram, featuring false endorsements from political figures or celebrities, which lead them to messaging platforms or fraudulent trading sites (Police warns of investment scams, 2025).

The increased vulnerability of younger people to online investment scams is a concern. The majority of investment scam victims were aged 30 to 49, making up 44.2 percent of victims of this scam type (SPF, 2024a). This age group's active pursuit of investment opportunities makes them prime targets for scammers, who use the allure of quick returns to persuade them into making risky investment decisions.

Job Scams and Singpass Credentials

Another recent phenomenon in Singapore's online scam landscape involves the compromising of Singpass (Singapore Personal Access). Singpass is a key digital identity system that provides access to government services and transactions. Scams involving Singpass have serious implications, as scammers can use it to commit identity theft, gain access to the victim's bank details, or make changes to official government records.

The most common method used by scammers is posting fraudulent job offers online or on communication platforms. In one recent case, some suspects were found to have allegedly sold their Singpass credentials for S$10,000 each, with the credentials then being used by scammers to open bank accounts and register for mobile phone lines (Koh, 2024). In other instances, scammers would send screenshots of Singpass QR codes and ask the victims to scan the code with their mobile phones, so that their personal information can be checked for the job (At least 219 victims duped, 2024). To address this growing problem, banks in Singapore will progressively implement Singpass face verification to strengthen authentication methods (MAS, 2024a).

Selling or giving away Singpass credentials to scammers is fundamentally different from falling for other scam types, as it is considered a criminal offense in Singapore. Those who willingly sell or give away Singpass credentials are legally liable for facilitating scam activities and will be treated as an accomplice rather than as a scam victim. The prevalence of scams involving Singpass exploits the public's digital trust in the government-linked digital identity system, indicating a lack of awareness of how their credentials can be misused.

Overconfidence in Identifying Scams

Users in Singapore exhibit overconfidence in identifying scams, which can be a significant risk in the digital age, where scams are becoming increasingly sophisticated. A survey by GASA (see Abraham et al., 2024) indicates that 62 per cent of survey respondents in Singapore are confident in identifying scams. Similarly, in a recent survey conducted by the Ministry of Digital Development and Information (MDDI), 56 percent of respondents across all age groups were moderately or extremely confident about identifying scam calls, while 45 percent felt the same about identifying scams on social media (Tan,

2024). However, 67 per cent of those aged 15 to 29 reported that they were either moderately or extremely confident in spotting scams on messaging platforms (Tan, 2024). However, the reality suggests that young people may not be as scam-savvy as they believe, with 29.7 per cent of scam victims aged 29 and below (SPF, 2024a). When the age group was expanded to include individuals under 50, 70.9 per cent of scam victims were found to be youths, young adults, and adults under 50 (SPF, 2024a).

People's confidence in their ability to spot online scams often do not match their actual success in detecting them (Wang et al., 2016). When many individuals, particularly those who spend much of their time online, believe they are capable of spotting fraudulent activity easily, this could lead to a false sense of security. Overconfidence can result in lower levels of vigilance, which causes people to overlook subtle red flags or dismiss warning signs. This may help explain why, despite high levels of digital literacy, users in Singapore remain vulnerable to falling for online scams.

Increased Vulnerability of Elderly Individuals

The increased vulnerability of elderly individuals is a significant concern. While the elderly represent one of the smaller age groups among scam victims in Singapore, the amount they lose per incident is significantly higher than that of victims in other age groups (SPF, 2024a). This is especially concerning because the financial losses sustained from such scams have the potential to deplete their life savings. Unlike younger people, they often lack the time and resources to rebuild their financial security after falling for a scam.

Many elderly individuals are susceptible to online scams due to their limited familiarity with digital platforms and heightened trust in strangers. This vulnerability may be in part attributed to their social isolation and loneliness (Wen et al., 2022). In one case, a 74-year-old man chatted online for six hours with a friendly roast duck seller he met on Facebook, who turned out to be a scammer who infected his phone with malware and syphoned funds from his online bank accounts (Sim, 2024). In another case, a 65-year-old Singaporean retiree lost her life savings of more than SG$1 million after falling for a scam by a Facebook friend (Hamzah, 2024). As online scams are becoming more sophisticated, the elderly are at heightened risk of falling for digital deception, which could have long-term impacts on their financial and emotional well-being if left unaddressed.

Misuse of AI in Scams in Singapore

The misuse of AI across the value chain has revolutionised the way scams targeting individuals in Singapore are conducted, making them more sophisticated and difficult to detect. AI is used in a variety of ways in scams targeting users in Singapore, ranging from chatbots that generate phishing emails at scale, to deepfake videos and voice cloning techniques used to impersonate trusted figures (Cyber Security Agency of Singapore, 2023; Chiu, 2024; Koh, 2023). Most Singaporeans have expressed a high level of awareness of the negative effects of AI technology on online scams, though this remains lower for AI-generated voices and videos (Abraham et al., 2024).

One observable trend in Singapore is the use of AI-powered chatbots like ChatGPT, which have facilitated the production of phishing emails and messages at scale. By using generative AI, messages have become more official-sounding with near-perfect language, mimicking genuine e-mails and messages from various organisations (Chia, 2024a). An example of this involved scammers posing as Consumers Association of Singapore (Case) officers distributing fake surveys via WhatsApp (Qing, 2024). Coupled with other tactics to make the scams look more credible, such as by using the https protocol and .com links, leveraging AI advancements make scam attempts harder to detect, as the language appears to be credible (Chia, 2024a).

Another notable trend is the increasingly sophisticated scams using deepfake technology. To illustrate the scale of the challenge, Singapore is facing an increase of 240 percent in deepfake attacks, the second highest in Asia-Pacific jointly with Cambodia (Koh, 2024). Scammers are leveraging AI to create realistic audio and video of trusted figures, such as government officials, in an attempt to deceive victims into believing that what they are seeing and hearing is genuine. Footage of Singapore's leaders Prime Minister Lawrence Wong and Senior Minister Lee Hsien-Loong were used to promote fraudulent investment products and circulated on social media platforms (Chiu, 2024; Koh, 2023). Deepfakes are also being used to create fake social media profiles, with a local Hong Kong syndicate found to have approached victims in Singapore with deepfaked images of good-looking women they found online (Ma, 2025).

There is emerging evidence that deepfakes are used in scams targeting individual users in Singapore. It is believed that these digital manipulations were used to change the appearances of scammers impersonating government officials or other high-ranking executives to persuade victims (MAS, 2025). Another risk involves the use of localised accents in deepfake audio, which can make scam calls appear far more authentic and difficult to detect. Victims in Singapore, who are used to specific local speech patterns, are more likely to trust a voice that sounds familiar, as opposed to distinctly foreign accents (Ng, 2025). The Cyber Security Agency of Singapore (2024, p.15) noted that the use of deepfake technology in online scams will continue to grow, "given the widespread accessibility of tools to create highly convincing deepfakes at a relatively low cost". That deepfake technology will be increasingly used in scams has caused widespread worry in Singapore, with more than three-quarters citizens and permanent residents expressing concern (Verian, 2024).

The value chain of AI-powered scams extends beyond Singapore, involving international networks of criminals from outside of the country, often based in other Southeast Asian countries or beyond (UNODC, 2024). These syndicates typically operate from countries with more lax regulatory environments or less stringent enforcement against cybercrime and extend their reach to Singapore by using digital platforms and untraceable payment methods to scam victims. Although online scams operate cross-border, digital governance fragmentation means that cybercrime laws are specific to local jurisdictions, making them difficult to enforce. Meanwhile, there is growing evidence of cross-border operations targeting users in Singapore. In January 2025, Hong Kong police arrested 31 individuals from a local syndicate involved in creating deepfake romance and investment scams to defraud victims in several countries, including Singapore (Ma, 2025). Syndicates capitalise on the transnational nature of these scams by combining social engineering tactics with AI-driven tools, tailoring their strategies to suit specific cultural and economic contexts. This targeted approach enhances the credibility of their scams and makes them seem more convincing.

POLICY GAPS IN ADDRESSING ONLINE FRAUD AND SCAMS

The Singaporean government has implemented a multi-layered strategy to combat online scams, focusing on prevention, enforcement, and public education. The Protection from Scams Bill, which was passed by the Singapore Parliament in January 2025 and expected to take effect in the second half of 2025, seeks to tackle the increasing number of self-effected transfers, in which victims willingly transfer the funds to scammers. To address this challenge, the bill empowers the police to issue Restriction Orders (ROs) to temporarily freeze the bank accounts of individuals suspected of falling victim to scams (Rajah & Tann Singapore, 2025). Despite critics of the Bill have argued that it may actively interfere with individuals' financial autonomy, the Bill is not expected to have much pushback from the public (Sun, 2024). A policy gap is that the Bill is relatively limited in scope, since the ROs currently do not cover other entities that are often involved in scam workflows such as cryptocurrency exchanges and e-wallet providers (Rajah & Tann Singapore, 2025; Sun, 2025).

Singapore is driving specific measures that digital platforms must adopt to prevent scam activities online. The Online Criminal Harms Act (OCHA), effective 1 February 2024, requires digital platforms such as Carousell and Facebook Marketplace to verify 'risky' sellers and advertisers against government-issued records (Lee & Tan, 2024). Facebook has since required all its advertisers to verify their identities by the end of June 2025, following a 12 per cent rise in scam ad reports during a pilot test from June to December 2024 (Chia, 2025). Carousell, on the other hand, was granted a six-month extension to reduce scams after showing an 11 per cent decline, but must verify all sellers by October if improvements stall (Chia, 2025). Meanwhile, platforms like Facebook, WhatsApp, Instagram, Telegram and WeChat are mandated to create a fast-track channel to receive and act on reports from the authorities (Lee & Tan, 2024). However, while the current law focuses on sellers and advertisers verification, scammers can also operate as buyers, creating a policy gap (Hamzah, 2024b). Another gap also exists when messaging apps like Instagram or Whatsapp, which are classified as online communication services, are used for informal e-commerce by scammers. In these scams, scammers can bypass government ID verification rules required for traditional ecommerce sites. Additionally, there is limited oversight if transactions are taken off-platform.

The government has also initiated efforts to enhance accountability among stakeholders in tackling online scams. Established by MAS, the Shared Responsibility Framework (SRF) aims to complement legislative efforts by distributing accountability for phishing scams between consumers, financial institutions, and telecommunication operators (telcos) (MAS, 2024b). However, the framework excludes scams where victims authorise payments to the scammer (i.e., self-effected transfers), as well as scams where victims were deceived into giving away credentials to the scammer directly. This means that scams with the highest total amount lost, such as investment scams and government official impersonation scams, are not covered by the SRF.

Singapore's current regulations do not explicitly address deepfakes, but there are several measures that can indirectly tackle this issue. The Protection from Online Falsehoods and Manipulation Act (POFMA) addresses the spread of falsehoods online and can be used to address manipulated content (Government of Singapore, 2025). Additionally, OCHA allows the government to direct digital platforms to remove potential scam related content, including those that are deepfake-enabled, to reach Singapore users (Ministry of Digital Development and Information, 2024). This is being complemented by efforts to strengthen the government's capabilities in addressing these threats, such as through industry collaboration and technological development of deepfakes detection by the SPF and the Home Team Science and Technology Agency (HTX) (Ministry of Digital Development and Information, 2024). This is being complemented by efforts to strengthen the government's capabilities in addressing these threats, such as through industry collaboration and technological development of deepfakes detection by the SPF and the Home Team Science and Technology Agency (HTX) (Ministry of Digital Development and Information, 2024). The government has stated that laws such as POFMA and OCHA can be used to address deepfakes in contexts outside of elections (Teo, 2024). Furthermore, the government has also publicly indicated its intent to introduce a Code of Practice on how to handle digitally manipulated content beyond election periods, suggesting a policy shift towards preventive measures in content governance (Rise of AI and deepfakes, 2025).

Beyond regulatory intervention, multi-stakeholder and cross-border collaborations are central to the Singaporean government's effort in tackling online scams. The SPF has established the Anti-Scam Command (ASCom) to coordinate efforts across various agencies to address scams in real time. The ASCom is a dedicated unit within the SPF that proactively detects and intervenes in potential scam situations by collaborating with banks and digital platforms. With staff from major banks co-located at the centre and leveraging advanced Robotic Process Automation technology to automate the process of information sharing and processing, the combined expertise of key stakeholders has the potential to

strengthen the public safeguards against online scams (SPF, 2024b). This success was demonstrated by the recovery of more than US$310 million from scammers between 2019 and 2023 (Wong, 2024a). Additionally, the ASCom also serves as a point of contact for cross-border collaborations. For example, ASCom and Malaysia's National Scam Response Centre (NRSC) successfully ran a joint anti-scam operation between February and March 2025, freezing more than 3400 bank accounts (Lim, 2025). While the ASCom has made significant strides in combating scams through its collaborative approach, there is still room for improvement, such as increasing the involvement of digital platforms involved in the scam workflow beyond Carousell and Shopee and enhancing cross-border cooperation with countries identified as scam hotspots (Chua, 2024).

Public education is another key pillar of Singapore's anti-scam efforts. Singapore has launched large-scale national anti-scam campaigns, such as the 'I can ACT against scams' campaign launched in January 2023. In the omni-channel campaign, anti-scam messages and advisories are widely disseminated across various channels, including television, radio advertisements, posters, digital ads, and local news outlets, typically with more publicity when there are emerging scam variants (Ministry of Home Affairs, 2024). Another notable innovation is the ScamShield app, which helps users to identify and block potential scam calls and messages. It also educates users about scam types and includes real-time alerts (Theseira, 2024). An important area to explore further is whether public education is backed by robust networks that can effectively reach and engage the most vulnerable communities, alongside the integration of comprehensive impact analysis.

POLICY RECOMMEDATIONS

Based on the identified gaps, the following are several recommendations that could improve the existing measures:

1. **Expand the scope of scam prevention:** The Protection from Scams Bill should explicitly include other entities such as cryptocurrency exchanges and e-wallet providers. This will ensure a more comprehensive framework for scam prevention across digital financial services.
2. **Enhance collaborative frameworks for combating online scams:** This could involve clarifying the roles and responsibilities of various stakeholders, including government bodies, digital platforms, and financial institutions, in addressing scam-related content and activities. Exploring incentives for proactive scam prevention measures across the digital ecosystem, potentially drawing on existing models like the SRF, could be beneficial. Additionally, the development and promotion of secure online transaction payment options that prioritise user safety are critical to maintaining a trustworthy digital ecosystem. Reviewing codes of practice for converging online communication and e-commerce services may also be necessary to establish clearer, shared expectations for mitigating risks. Furthermore, fostering strengthened actions by law enforcement, alongside supportive multi-stakeholder cooperation, is crucial to tackling the organized criminal networks as the root cause, which are understood to be a primary source of the issue.
3. **Include deepfakes in broader online harm prevention:** Given that Singapore already has a robust legal framework to address online harms, strengthening and expanding existing laws could be a viable alternative to creating standalone deepfake regulations. One way to do so can be by enhancing the Code of Practice for Online Communication Services to include specific provisions for detection and labelling.
4. **Implement active deepfake detection and traceability measures:** Digital platforms should be expected to develop and adopt AI-driven tools for deepfake detection and proactively labelling them when they appear on the platform. In particular, private messaging apps like WhatsApp would benefit from centralised content scanning system that can actively flag scam-related

deepfake content. Additionally, tracing the origin, alterations, and distribution of content can help curb the spread of deepfakes at their source.

5. **Strengthen informal networks as support for vulnerable groups:** Data has shown that the elderly are at risk of losing large amounts of funds - including their life savings - to online scams, while younger people are increasingly more vulnerable to investment scams. To help mitigate these risks, informal networks such as family members, close friends and community groups or leaders can play a crucial role in strengthening community bonds, sharing best practices for anti-scam prevention, and providing support.
6. **Conduct comprehensive impact analysis:** Impact analysis can assess how well interventions have reduced users' vulnerability to social engineering tactics, which underlie much of the self-affected transfers. Understanding the effectiveness of existing strategies can reveal areas where users are still susceptible to manipulation and help design more targeted approaches.
7. **Improve the participation of digital platforms at Anti-Scam Command:** Requiring digital platforms to engage with the Anti-Scam Command as a proportionate response when platforms are found to be behaving irresponsibly would foster a more coordinated approach to tackling scams in Singapore. This targeted strategy allocates resources and oversight where they are needed by focusing on platforms that have demonstrated a need for improved scam prevention outcomes.
8. **Strengthen regional cooperation:** Scam syndicates operating in other Southeast Asian countries like Cambodia or Myanmar can be difficult to track and prosecute due to differences in governance, as well as varied levels in law enforcement and cybersecurity capabilities. Greater regional cooperation through ASEAN could help standardise approaches and improve unified responses to cross-border online scams across Southeast Asia. This should include standards for criminalising the use of deepfakes for malicious purposes, such as scams.

REFERENCES

At least 219 victims duped into revealing Singpass credentials to scammers since January. (2024, May 19). *The Straits Times.* https://www.straitstimes.com/singapore/at-least-219-victims-duped-into-revealing-singpass-credentials-to-scammers-since-january

Abraham, J., Rogers, S., Njoki, C., & Greening, J. (2024). *The State of Scams in Singapore 2024.* Global Anti-Scam Alliance. https://www.gasa.org/_files/ugd/7bdaac_7dceee4e90f9493eac18bccb1425304f.pdf

Chia, O. (2024a, July 30). 13% of phishing scams analysed likely to be AI-generated: CSA. *The Straits Times*. https://www.straitstimes.com/singapore/13-of-phishing-scams-analysed-likely-to-be-ai-generated-csa

Chia, O. (2024b, October 30). Look out for these scams on Facebook, Instagram and WhatsApp. *The Straits Times*. https://www.straitstimes.com/singapore/look-out-for-these-scams-on-facebook-instagram-and-whatsapp

Chia, O. (2025, March 11). Facebook advertisers must verify their identities by end-June following rise of scam ads. The Straits Times. https://www.straitstimes.com/singapore/all-facebook-advertisers-need-to-verify-identity-by-june-following-rise-of-scam-ads

Chiu, C. (2024, January 4). PM Lee warns against responding to deepfake videos of him promoting investment scams. *The Straits Times*. https://www.straitstimes.com/singapore/pm-lee-warns-against-responding-to-deepfake-videos-of-him-promoting-investment-scams

Chua, N. (2024, May 9). MHA to consider mandating deployment of staff from online platforms to police's Anti-Scam Command. *The Straits Times*. https://www.straitstimes.com/singapore/politics/mha-to-consider-mandating-deployment-of-staff-from-online-platforms-to-police-s-anti-scam-command

Cyber Security Agency of Singapore. (2024). *Singapore Cyber Landscape 2023*. https://www.csa.gov.sg/resources/publications/singapore-cyber-landscape-2023/

Government of Singapore. (2025, March 20). *Singapore's fight against Misinformation.* https://www.gov.sg/explainers/singapore-fight-against-misinformation

Hamzah, A. (2024, November 13). Over $1m lost in 15 days: S'porean retiree loses life savings in scam by fake Facebook friend. *The Straits Times*. https://www.straitstimes.com/singapore/gone-in-15-days-40-years-of-savings-amounting-to-more-than-1-million

Hamzah, A. (2024, November 25). At least 877 people duped by fake buyers on Carousell since December. *The Straits Times.* https://www.straitstimes.com/singapore/at-least-877-people-duped-by-fake-buyers-on-carousell-since-december

Digital Society. (n.d.). Infocomm Media Development Authority. https://www.imda.gov.sg/About-IMDA/Research-and-Statistics/Digital-Society

Koh, S. (2023, December 29). Deepfake video of DPM Lawrence Wong promoting investment scam circulating on social media. *The Straits Times*. https://www.straitstimes.com/singapore/deepfake-video-of-dpm-lawrence-wong-promoting-investment-scam-circulating-on-social-media

Koh, S. (2024, April 23). 78 people probed for allegedly giving Singpass details to scammers. *The Straits Times.* https://www.straitstimes.com/singapore/courts-crime/78-people-investigated-for-allegedly-giving-singpass-details-to-scammers

Koh, F. (2024, 22 November). Singapore registers Asia-Pacific's biggest spike in identity fraud, driven by deepfake surge. https://www.channelnewsasia.com/singapore/identity-fraud-deepfakes-scams-ai-4761836

Lee, L. Y. & Tan, C. (2024, June 29). New codes of practice require Carousell, Facebook to verify 'risky' sellers, advertisers to curb scams. *The Straits Times*. https://www.straitstimes.com/singapore/new-codes-of-practice-require-carousell-facebook-to-verify-risky-sellers-advertisers-to-curb-scams

Lim, K. (2025, March 19). More than 850 people being investigated in Singapore-Malaysia joint anti-scam operation; victims lost at least $8.1m. A*siaOne.* https://www.asiaone.com/singapore/more-850-people-being-investigated-singapore-malaysia-joint-anti-scam-operation-victims

Ma, J. (2025, January 5). Hong Kong police arrest 31 over deepfakes used to scam victims in Singapore, Malaysia. *The Straits Times*. https://www.scmp.com/news/hong-kong/law-and-crime/article/3293476/hong-kong-police-arrest-31-who-used-deepfakes-scam-victims-singapore-malaysia

Ministry of Digital Development and Information. (2024, February 24). *MCI's response to PQ on Regulations to Tackle Deepfake Software Used in Scam and Fraud Cases*. https://www.mddi.gov.sg/media-centre/parliamentary-questions/pq-on-regulations-to-tackle-deepfake-software/

Ministry of Home Affairs. (2024, January 9). *Written Reply to Parliamentary Question on Frequency of Anti-Scam Campaigns in the Media.* http://mha.gov.sg/mediaroom/parliamentary/written-reply-to-pq-on-frequency-of-anti-scam-campaigns-in-the-media/

Monetary Authority of Singapore. (2024a, September 18). *Major retail banks to introduce Singpass Face Verification, further strengthening resilience against phishing scams* [Press release]. https://www.mas.gov.sg/news/media-releases/2024/major-retail-banks-to-introduce-singpass-face-verification

Monetary Authority of Singapore. (2024b, October 24). *MAS and IMDA Announce Implementation of Shared Responsibility Framework from 16 December 2024* [Press release]. https://www.mas.gov.sg/news/media-releases/2024/mas-and-imda-announce-implementation-of-shared-responsibility-framework-from-16-december-2024

Monetary Authority of Singapore. (2024c, November 30). *Joint Advisory on Rise of Government Official Impersonation Scam Variant Featuring Impersonation of Banks* [Press release]. https://www.mas.gov.sg/news/media-releases/2024/rise-in-government-official-impersonation-scam-variant

Monetary Authority of Singapore. (2025, March 12). *Joint Advisory on Scams Involving Digital Manipulation* [Press release]. https://www.mas.gov.sg/news/media-releases/2025/joint-pnr-by-spf-mas-and-csa

Ng, D. (2025, March 19). Commentary: A close call showed me that anyone can get scammed – even me. *Channel News Asia.* https://www.channelnewsasia.com/commentary/scam-phone-call-fraud-prevention-awareness-4997496

Police warns of investment scams; at least $32.6 million lost in over a month. (2025, February 11). *CNA.* https://www.channelnewsasia.com/singapore/investment-scams-victims-social-media-dating-app-coffee-meets-bagel-4930741

Principal. (2024). 2024 Global Financial Inclusion Index. https://www.principal.com/financial-inclusion

Qing, A. (2024, November 22). Shopping survey on WhatsApp that offers $13 payment is a new scam, warns Case. *The Straits Times*. https://www.straitstimes.com/singapore/case-warns-against-scam-on-whatsapp-which-promises-13-for-completing-fake-survey

Rajah & Tann Singapore. (2025, February 7). *Protection from Scams Bill Passed in Parliament.* https://sg.rajahtannasia.com/viewpoints/protection-from-scams-bill-passed-in-parliament/

Rise of AI and deepfakes in lead-up to GE2025. (2025, April 11). Channel News Asia. https://www.channelnewsasia.com/interactive/ge2025-deepfake/

Sim, S. (2024, November 15). 74-year old man loses $70k after downloading third-party app to buy Peking duck. *The Straits Times*. https://www.straitstimes.com/singapore/74-year-old-man-loses-70k-after-downloading-third-party-app-to-buy-roast-duck

Singapore Police Force. (2024a). *Annual Scams and Cybercrime Brief 2024.* https://www.police.gov.sg/Media-Room/Police-Life/2025/02/Five-Things-You-Should-Know-about-the-Annual-Scams-and-Cybercrime-Brief-2024

Singapore Police Force. (2024b, September 6). *Joint operation between the Anti-Scam Centre and six partnering banks led to the disruption of more than 2,016 scams* [Press release]. https://www.police.gov.sg/media-room/news/20240906_joint_operation_between_the_anti_scam_centre

Sun, D. (2024, November 17). How did S'pore end up needing 'nanny' laws to save scam victims from themselves? *The Straits Times*. https://www.straitstimes.com/singapore/how-did-we-end-up-needing-nanny-laws-to-save-scam-victims-from-themselves

Sun, X. (2025, January 7). *Second Reading of the Protection From Scams Bill - Wrap-Up Speech by Ms Sun Xueling, Minister of State, Ministry of Home Affairs and Ministry of Social and Family Development* [Transcript]. Ministry of Home Affairs. https://www.mha.gov.sg/mediaroom/parliamentary/second-reading-of-the-protection-from-scams-bill-wrap-up-speech/

Tan, C. (2024, November 14). Young people say in MCI survey say they can identify scams, but police crime statistics show otherwise. *The Straits Times*. https://www.straitstimes.com/singapore/courts-crime/young-people-say-in-mci-survey-they-can-identify-scams-but-police-crime-statistics-show-otherwise

Tan, C. (2025, March 19). Woman loses $1.2m to scammers who pretended to be officers from police's Anti-Scam Centre. https://www.straitstimes.com/singapore/woman-loses-1-2m-to-scammers-who-pretended-to-be-officers-from-polices-anti-scam-centre

Technode Global. (2025, February 11). *Digital 2025: Nearly Two Thirds of Southeast Asia's Population are on Social Media.* https://technode.global/2025/02/11/digital-2025-nearly-two-thirds-of-southeast-asias-population-are-on-social-media/

Teo, J. (2024, October 15). Closing Speech by Minister Josephine Teo at the Second Reading of the ELIONA Bill. Ministry of Digital Development and Information. https://www.mddi.gov.sg/closing-speech-by-minister-josephine-teo-at-the-second-reading-of-the-eliona-bill/

Theseira, J. (2024, September 29). Easier to check, report and recall: How the new ScamShield Suite can help you outsmart scammers. *The Straits Times*. https://www.straitstimes.com/singapore/easier-to-check-report-and-recall-how-the-new-scamshield-suite-can-help-you-outsmart-scammers

United Nations Office on Drugs and Crime (UNODC). (2024). *Transnational Organised Crime and the Convergence of Cyber-Enabled Fraud, Underground Banking and Technological Innovation in Southeast Asia: A Shifting Threat Landscape.* https://www.unodc.org/roseap/uploads/documents/Publications/2024/TOC_Convergence_Report_2024.pdf

Verian. (2024, April 8). *Three quarters of Singaporeans concerned about the use of deepfakes in scams.* https://www.veriangroup.com/news-and-insights/three-quarters-of-singaporeans-concerned-about-the-use-of-deepfakes-in-scams

Wang, J., Li, Y., & Rao, H. R. (2016). Overconfidence in Phishing Email Detection. *Journal of the Association for Information Systems, 17*(11), 759-783. https://doi.org/10.17705/1jais.00442

Wen, J., Yang, H., Zhang, Q., & Shao, J. (2022). Understanding the mechanisms underlying the effects of loneliness on vulnerability to fraud among older adults. *Journal of Elder Abuse and Neglect, 34*(1), 1-19. https://doi.org/10.1080/08946566.2021.2024105

Wong, S. (2024a, August 4). More countries set up anti-scam centres: increased teamwork, speed vital to retrieve lost funds: SPF. https://www.straitstimes.com/singapore/more-countries-set-up-anti-scam-centres-increased-teamwork-speed-vital-to-retrieve-lost-funds-spf

Wong, S. (2024b, November 13). $1.4 trillion lost to scams globally: S'pore victims lost the most on average: Study. *The Straits Times*. https://www.straitstimes.com/world/14-trillion-lost-to-scams-globally-s-pore-victims-lost-the-most-on-average-study

Yasmine, R. (2025, February 9). Over $1.32 million seized, 46 people arrested in operation targeting higher-loss scams. *The Straits Times*. https://www.straitstimes.com/singapore/over-1-32m-seized-46-people-arrested-in-operation-targeting-higher-loss-scams

Online Fraud and Scams in South Korea

Rosa (Hyun Kyong) Lee

8

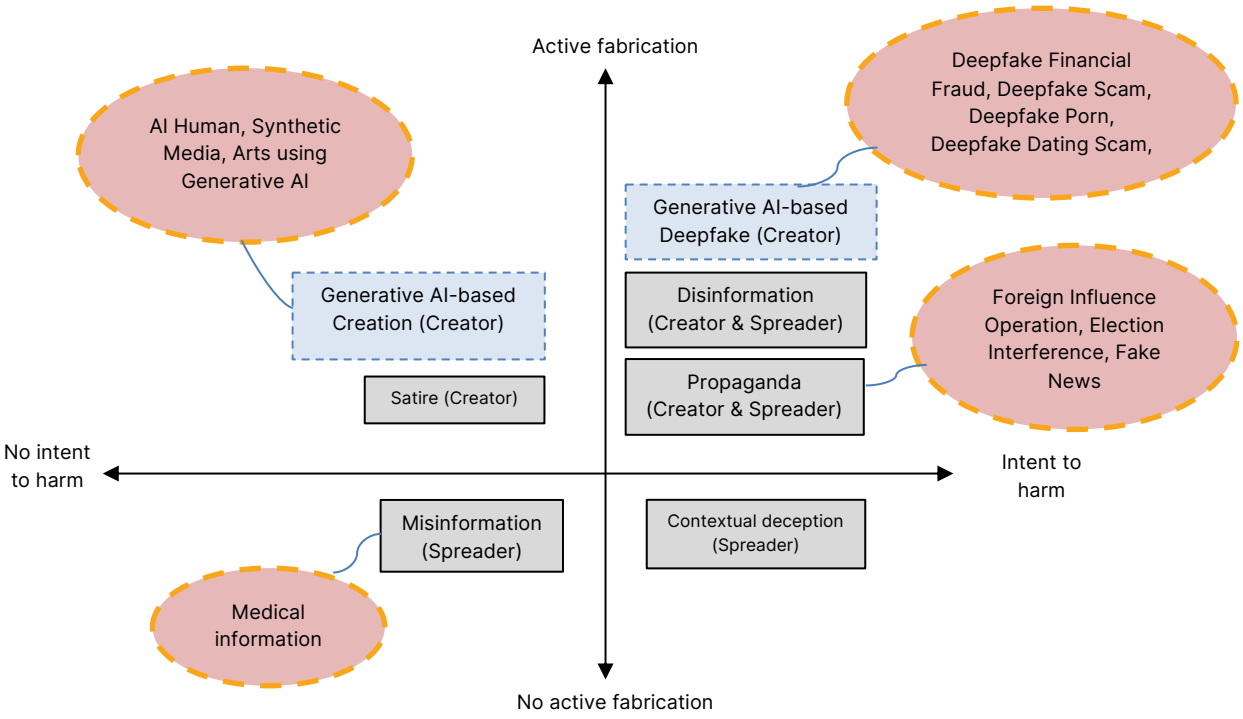
Online Fraud and Scams in South Korea

Rosa (Hyun Kyong) Lee[20]

INTRODUCTION

Recent advancements in generative AI technology have made the production of disinformation and online scams more sophisticated and complex. While traditional disinformation and online fraud was primarily created using simple digital tools (or no digital tools), the emergence of generative AI has revolutionized this process by leveraging big data and advanced algorithms to automate and rapidly generate false contents. Most literature regarding the reliability of information focuses on differentiating misinformation and disinformation. As illustrated in the below figure, Lesher et al. (2022) distinguished between these concepts based on two key factors: whether there is an intent to cause harm to others or society and whether the information is actively manipulated. In the figure, those rectangles drawn with a solid line represent the original framework proposed by Lesher et al.(2022); disinformation involves both intentional harm and active fabrication (Lesher et al., 2022).

Figure 8.1 A Typology of Untruths Online and Emerging Threats



Source: Author's modification from Lesher et al. (2022)

With the emergence of generative AI technology, new threats are emerging. Before the era of generative AI, the boundaries of each quadrant were relatively clear and organizational resources and strategies could be unified to address each quadrant. However, these boundaries are becoming blurry as new technology comes in. The rectangles with a dashed line in Figure 1 represent new threats stemming from generative AI (The dashed circle shows examples of each concept.). To identify such new threats, many expressions are utilized interchangeably, such as fake news, misinformation, disinformation, online propaganda, synthetic media, deepfake, and online scams.

However, those threats require distinctive responses depending on the domain of the threats and the malicious use of technology. For example, organizations used to deal with traditional financial fraud are

[20] Associate Research Fellow, Korea Information Society Development Institute

now required to respond to financial fraud using AI technology. Existing laws and policy responses are not yet equipped to fully react to the misuse of AI technology with active fabrication.

This paper explores how emerging threats, particularly those involving generative AI, are infiltrating Korean society and examines how current policies respond to these challenges. It begins by outlining recent trends of online scams in Korea, including the misuse and application of generative AI in such scams. It then reviews existing national policies and cross-border strategies aimed at addressing online scams and fraud using AI technology. The chapter also assesses the potential societal implications of AI-generated online scams and fraud for the Korean population. In addition, it identifies the key stakeholders in South Korea involved in addressing AI-generated online scams and fraud. Finally, it highlights best practices and lessons learned from Korea's experience.

PATTERNS AND TRENDS

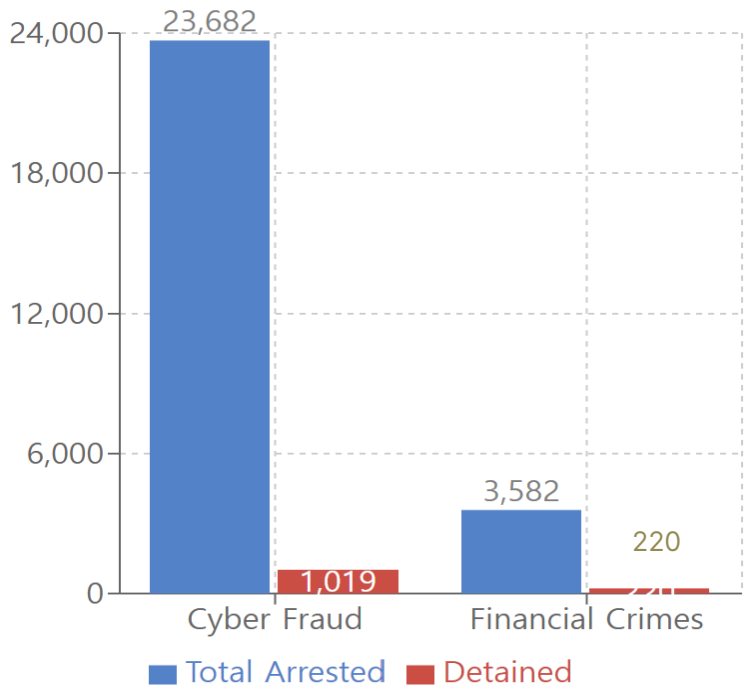
South Korea has experienced a rise in sophisticated financial and online fraud cases leveraging generative AI and deepfake technology. The increasing use of generative AI threatens traditional policy measures and prevention approaches. Online scams evolved to adopt generative AI technology and created technological and legal gaps to detect, identify the criminals, punish the victims, and protect users. These crimes impact on Korean society, especially on economic damage, privacy, and social trust, and demands public attention and policy response to make Korean society prepared for the era of general artificial intelligence.

In South Korea, cyber scam activities have been a long-time social problem, and patterns and trends have increasingly become sophisticated with the rapid development of generative AI. According to the Ministry of Science and ICT(MSIT), cyber fraud is one of the three cyber threat cases in 2024 (Ministry of Science and ICT, 2024), with software supply chain attacks and advanced ransomware attacks. Common type of cyber scam activities in Korea include 1) impersonation of public institutions (messages pretending to be from public institutions about tax refunds, fines, etc.), 2) holiday gifts scams(online transfers or gift certificates), 3) Non face-to-face transactions(scams involving delayed delivery, out-of-stock items or used items), 4) fake online stores(fraudulent online shopping sites) and reviews. This type of cyber scam may not necessarily utilize generative AI but might require some level of digital competency or traditional computer skills.

Major fraud cases utilizing AI technology could entail AI-enhanced voice phishing attacks, deepfake-powered fraud schemes, deepvoice technology exploitations. For example, criminals use ChatGPT (or other similar generative AI tools) to craft highly personalized phishing scenarios tailored to victims' specific vulnerabilities. Another notorious example is kidnapping fraud using synthetic videos. The National Investigation Headquarters of the Korean National Police Agency reported cases where criminals used deepfake technology to create videos appearing to show kidnapped children, demanding ransom from parents (November 7, 2024). Deepvoice technology could be exploited for family voice cloning and corporate command chain manipulation.

Currently, published statistics regarding cyber fraud do not necessarily differentiate AI-based cyber fraud and traditional cyber fraud. Some statistics offer a bird eye view of how much cyber fraud Korean society is dealing with. According to the Korean National Policy Agency (2023), a total of 27,264 individuals were apprehended for cyber fraud (23,682 suspects were arrested, with only 1,019 of them detained) and financial crimes (3,582 suspects were arrested, with 220 of them detained) during an eight-month nationwide crackdown in 2023. It is unknown what percentage of these cyber frauds constitute scams using generative AI technology.

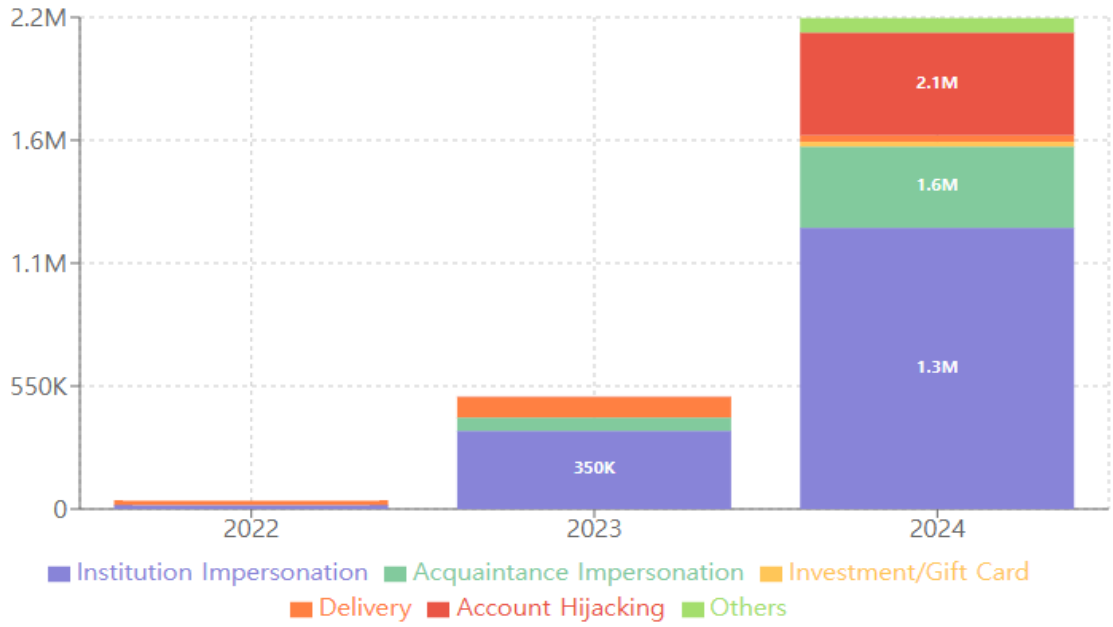
Figure 8.2 Cyber fraud and financial crime arrests in 2023



Source: Korean National Policy Agency (2023)

Meanwhile, the most common type of cyber scam in Korea is using text messages. According to the Ministry of Science and ICT and the Korea Internet & Security Agency (KISA), text scams in Korea typically involve institution impersonation, acquaintance impersonation, investment and gift certificate fraud, delivery fraud, and other types of scams. According to statistics from relevant authorities on text scams from 2022 to 2024, the most common type involved impersonating public institutions, accounting for 1.62 million cases (59%) (Financial Services Commission, January 20, 2025). Messages impersonating acquaintances, such as wedding invitations or funeral announcements, amounted to 423,191 cases (15.1%). Additionally, in 2024, there was a significant rise in messages impersonating investment opportunities (stocks and cryptocurrencies) or offering gift certificates, with approximately 21,088 cases (1.0%). Similarly, there was a sharp increase in account hijacking types compared to the previous year, from 2,315 cases (0.5%) in 2023 to 459,707 cases (20.9%) in 2024.

Figure 8.3 Total Cases of Fraud in South Korea by Type



Source: Financial Services Commission (2025)

Public awareness of cyber scams using generative AI is well-known in Korea due to a notorious fake investment scheme involving celebrities. In 2022, scammers produced deepfake videos of two top film stars to promote a bogus investment opportunity (YTN, 2024). Victims, trusting the familiar faces of top stars, handed over their assets. When the fraud came to light, it tested how Korean law addresses AI manipulation in the context of scams. If the perpetrators were caught, they would face traditional fraud charges; the deepfake aspect is an aggravating factor but not a separate offense in Korea.

The case highlighted a legal gray area: using another person's likeness in advertising without consent is typically a civil issue (right of publicity), but the incident here was part of criminal fraud. Victims were deceived not only by false promises but also by identity misuse (Baek & Lee, 2024). Beyond fraud charges, the actors/actresses whose faces were used could potentially sue for misuse of their image. However, if the criminals are overseas or their identities are unknown, neither the criminal charge nor civil remedies are effective. Perhaps governments need better platforms to screen for fake celebrity endorsements more rigorously to protect the public.

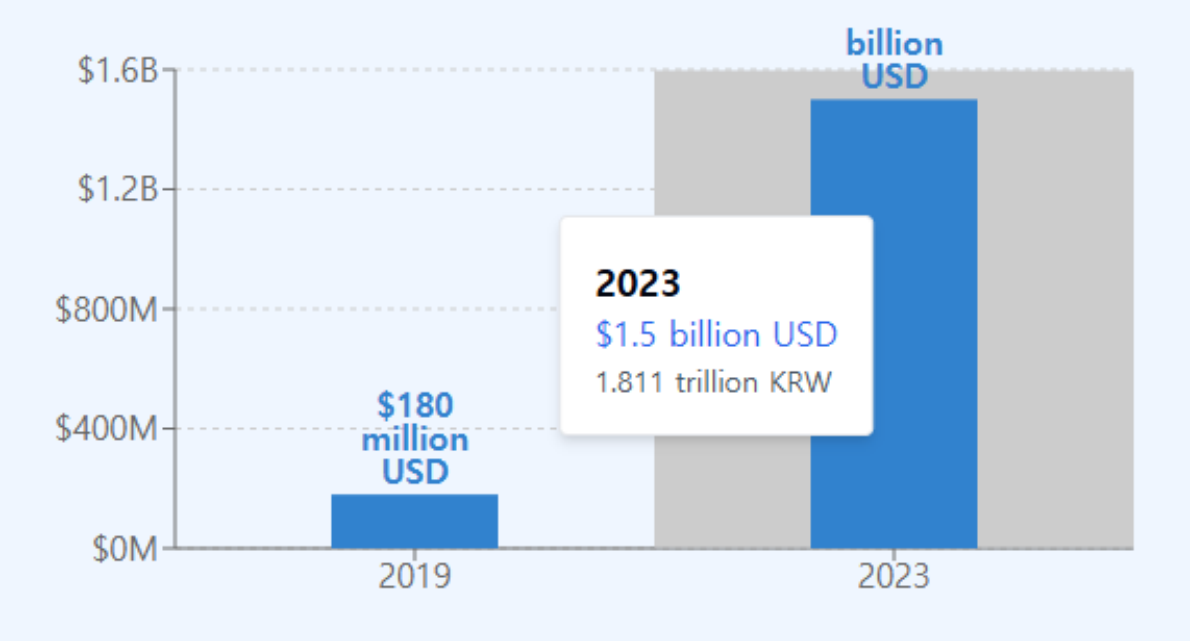
These patterns and trends highlight a critical point in South Korea's battle against cyber fraud and the emerging threats posed by the misuse of AI technology. As generative AI technologies become more accessible and sophisticated, the line between traditional cyber scams and AI-enhanced fraud continues to blur, creating challenges for statistical tracking and legal frameworks. Against this backdrop, the following section examines existing national policies and cross-border strategies in Korea.

THE POTENTIAL IMPLICATIONS OF ONLINE SCAMS AND FRAUD

Cyber fraud, whether enabled by the latest technology like generative AI, affects society both economically and psychologically, impacting both victims and society as a whole. Public trust in technology can be weakened, thereby hampering the vitality of the innovation ecosystem. However, it can also facilitate safety-first development, leading to some positive changes.

The financial impact of online scams has increased significantly in South Korea between 2019 and 2023. According to the National Assembly's audit data submitted by the Korean National Policy Agency, the total cost of cyber fraud reached 1.811 trillion KRW (approximately $1.5 billion) in 2023 (Baek, September 9, 2024). Over the past five years, the total cost of cyber fraud has increased eightfold, from 222.2 billion KRW (approximately 180 million USD) in 2019 to 1.811 trillion KRW in 2023. While the number of cyber fraud cases has increased, the detection rate of cyber fraud cases has decreased over the years, from 77.6% in 2019 to 58% in 2023.

Figure 8.4 Total Cost of Cyber Fraud in 2019 and 2023 in USD



Source: Baek (2024)

The economic cost of online scams includes damage and destruction of data, stolen money, lost productivity, theft of intellectual property, and other related expenses. Even in the case of online financial fraud, the damage to individuals and organizations is not just economic. The emotional distress that the victims experience cannot be quantified easily. Some victims even committed suicide in Korea, and other victims experienced psychological damage.

The rise of online scams poses a threat to the broader digital ecosystem by eroding public trust. Users may limit their online activities or opt out of particular platforms altogether. This could lead the public to become reluctant to adopt new technologies or services, particularly among vulnerable populations. There are certain groups in Korea that are more vulnerable to online scams: foreign-born populations and immigrants may experience language difficulties, leaving them susceptible to voice phishing. It is likely that this group has never had such experience nor had social networks that could warn of the possibility of online scams.

The public's trust in online platforms could be diminished if the platform fails to respond effectively to online scam incidents and does not adequately protect its users. The success of the platform economy depends on users' trust in the safety of using the digital platforms. Thus, platform companies redirect their operational resources to fraud prevention and resolution. Although this could enhance the overall safety of the digital ecosystem in Korea, it could also put a significant burden on content moderation. This could lead to a digital divide among companies, with only large tech platforms that have sufficient

resources likely to survive, while small companies or startups may struggle in the long run due to inadequate security measures.

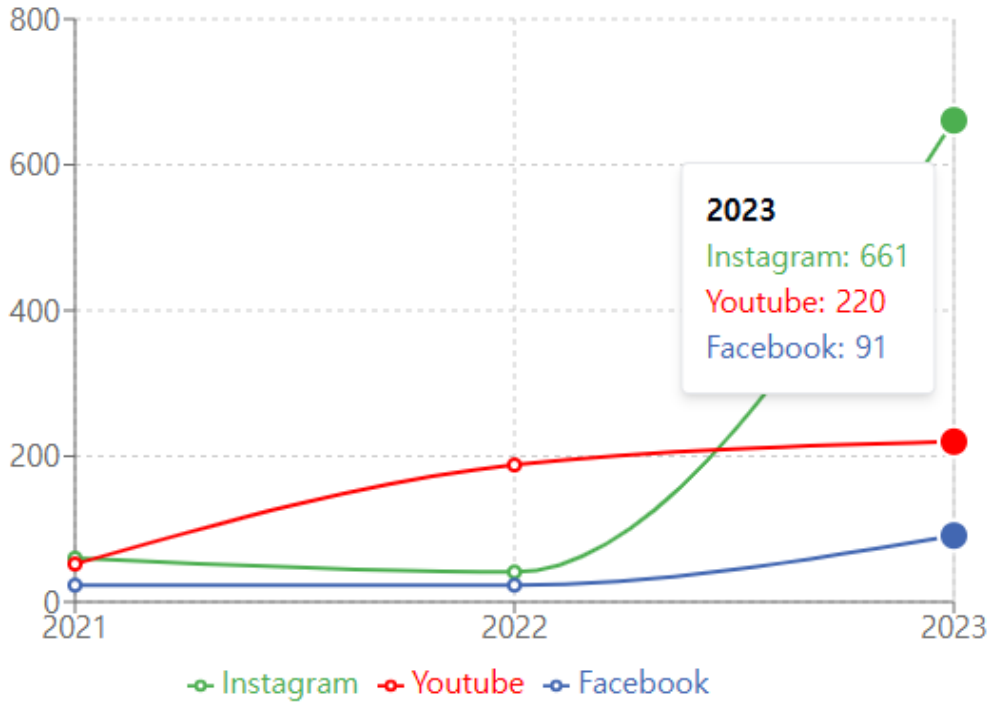
In this way, online scams around the world definitely push for security-first development. Financial sectors will increase their cyber security efforts, but malicious actors almost always find a way to circumvent or attack the secured network. With the development of AI-technology, it is getting easier for malicious actors to implement their plan, even without deep coding knowledge or computer background.

While the technical challenges of combating online fraud continue to grow, societal awareness represents another critical line of defense. In this regard, South Korea demonstrates some encouraging trends. Many celebrities openly share their experience of being scammed (not necessarily online scams), and voice phishing is often used as a subject for satire and comedy. According to a survey conducted by KBS 1TV, 94.7% of adults in South Korea reported having heard of or being familiar with electronic financial fraud crimes such as voice phishing, smishing, pharming, and messenger phishing (KBS, 2022). Approximately 35.2% of the respondents reported being very familiar with these fraud schemes, while 59.5% had general awareness, and only 0.1% of respondents stated that they were entirely unaware of them. Additionally, 86.5% of respondents reported having received fraudulent messages or calls at least once, and 78.7% believe they can become a victim at any time, demonstrating the prevalence of online fraud in South Korea.

However, specific populations are still vulnerable to this type of scam. According to Korea Research's survey result, among the elderly over 60 years of age, voice phishing and impersonation crimes using information and communication networks are increasing, and more than 80% of the elderly are concerned about crimes using information and communication networks (Korea Research, 2024).

Meanwhile, younger generations may be more susceptible to online shopping fraud due to their increased reliance on online shopping. According to the Korea Consumer Agency (KCA), complaints related to fraudulent overseas e-commerce sites increased from 251 cases in 2021 to 1,372 cases in 2023 (Sung, 2025). In particular, seven out of 10 incidents occurred at shopping malls accessed while viewing Instagram (41.8%), YouTube (25.3%), or Facebook (7.5%) (KCA, 2025). The most common fraud tactic is 'brand impersonation,' where consumers are led to process payments but never receive the goods (47.1%) (KCA, 2025). Another common tactic was delivering counterfeit or low-quality items instead of what was advertised (46.5%) (KCA, 2025).

Figure 8.5 Medium of Online Shopping Fraud in Korea



Source: Sung (2025)

The multifaceted impact of cyber fraud in South Korea reveals a complex landscape for policymakers. As the financial toll continues to rise dramatically, the decreasing detection rate signals a concerning trend that demands more sophisticated countermeasures. Beyond the quantifiable economic damage lies a more profound societal impact: eroding trust in digital platforms, psychological trauma for victims, and the potential for a bifurcated digital ecosystem where only resource-rich companies can afford security measures for AI-enabled financial fraud.

POLICY OVERVIEW: STAKEHOLDERS MAPPING AND NATIONAL POLICIES IN ADDRESSING SCAMS

The Role of Stakeholders in Addressing Scams

Financial fraud cases involving generative AI and deepfake technology in South Korea are on the rise, prompting various institutions to take action. Firstly, the Office of Government Policy Coordination (OGPC) is leading a whole-of-government task force to combat and eradicate online scams and other related crimes. In 2021, OGPC, in collaboration with the Ministry of Science and ICT (MSIT), the Korea Communications Commission (KCC), and the National Policy Agency (NPA), launched the “Whole-of-Government Task Force on Telecommunication Financial Fraud Response” (KISA, 2024). However, the scope of the OGPC covers all political affairs in the country under the Prime Minister’s Secretariat. Establishing a dedicated task force to address digital crimes and AI-based threats may be a more effective approach for combating future crimes.

Depending on the nature of the scam activities, there are many different organizations for anti-scam initiatives in South Korea. If the scam involves cyber fraud or cyber financial fraud, the Korean National Police Agency, Financial Services Commission, and Financial Supervisory Service are responsible for

investigation, support measures, and reporting. For public awareness campaigns and other educational activities, the Ministry of Science and ICT, along with the Ministry of Interior and Safety, takes the lead. Here are some lists of institutions in Korea for anti-scam initiatives.

- Korean National Policy Agency
- The Ministry of Science and ICT
- The Ministry of Interior and Safety
- Financial Service Commission (FSC)
- Financial Supervisory Service (FSS)
- Financial Security Institute (FSI)
- Korea Institute of Finance (KIF)
- Korea Internet & Security Agency (KISA)
- Korea Financial Crime Prevention Association (KFCPA)

These organizations have implemented diverse strategies and are fostering international cooperation to address the growing threat of sophisticated scam operations. Strategies include preparing technological response and reporting systems, public awareness and education, and legislative development; their initiatives work separately depending on the jurisdiction of the corresponding organizations. The role of the National Police Agency for preventing and combating online scams is especially effective when there are organized fraud enablers – utilizing their 'Cyber Crime Reporting System'.

The role of the Financial Service Commission (FSC) and Financial Supervisory Service (FSS) is also essential when the fraud is related to a financial scam. FSC is a government agency with statutory authority over financial policy and regulatory supervision. FSS is a specially legislated quasi-government supervisory authority and charged with financial supervision across the entire financial sector. In terms of anti-scam activities, their primary role is consumer protection and preventing voice phishing.

Government branches, such as the Ministry of Science and ICT and the Ministry of Interior and Safety, typically coordinate their policies with the OGPC by allocating budgets for public awareness programs, campaigns, and other supportive initiatives.

Alongside the government branches, there are other types of organizations for promoting anti-scam related initiatives and researching relevant topics. For example, KISA is an organization promoting internet and information security, founded in 2009. KISA operates 'Boho Nara & KrCERT/CC' to countermeasure hacking and virus attacks, developing technical responses to attack tools. For individuals, KISA's 'Boho Nara & KrCERT/CC' offers smishing and quishing verification services. Their service also targets corporations and entities, as well as small and medium-sized enterprises (SMEs). Another example is the Korea Institute of Finance. KIF leads research to advance the financial industry and facilitate the realization of the 'Information Age' across the financial sector.

Lastly, the private sector is taking an essential role by promoting communal benefits through associations. For example, the Korea Financial Crime Prevention Association (KFCPA) was established to research and counteract serious financial crimes, promoting awareness of the risks and effective prevention methods to the public in order to prevent the spread of damage.

However, the current status lacks control towers, thus making these anti-scam related efforts fragmented. Many different agencies have their own jurisdictions, and there is a possibility of overlapping responsibilities in tackling scam-related activities. Since these organizations possess their own expertise and resources and attempt to prepare for future threats enabled by AI technology, there may be overlapping policy responses to generative artificial intelligence for financial fraud. Without an adequate

control tower or regulatory governance, it would be challenging to address the ever-increasing complexity of generative AI-based online scams.

The fragmentation of authority among these stakeholders becomes problematic when addressing complex issues such as data privacy and governance, which are crucial to combating AI-generated scams. The current state of data and privacy governance in South Korea can illustrate the challenging situations related to online scams. First, AI-generated scams and deepfakes raise serious privacy concerns. They often involve the unauthorized use of a person's image, voice, or other personal data, directly conflicting with principles of personal data protection. South Korea's Personal Information Protection Act (PIPA) is the primary law safeguarding personal data, but it faces limitations. PIPA generally applies to organizations or businesses handling personal data. If a private individual creates a deepfake of someone using photos scraped online, that act might not fall under PIPA's enforcement provisions.

As new types of identity theft, such as pretending to be someone else on social networking services like Facebook and Instagram, as well as dating apps, have been increasing, new data regulations may be necessary. This type of identity theft does not involve traditional personal information, such as resident registration numbers, nor is it committed for monetary gain. A photo or video of a person is "personal data", and using it in a deepfake could be seen as unauthorized processing of that data.

However, in essence, data governance itself does not suffice to protect the public against online scams. Although South Korea classifies biometric data (face images, voiceprints) as sensitive information protected by strict consent requirements, enforcement is challenging when data is scraped from public social media or when the AI service is located overseas. Thus, new data governance should encompass a global scope so that international cooperation is possible if overseas enforcement is not possible.

The institutional framework addressing AI-enabled financial fraud in South Korea reflects significant challenges in governance. While individual agencies, such as the National Police Agency, FSC, and KISA, have developed specialized competencies within their respective domains, the absence of a dedicated control tower creates coordination gaps. Moving forward, South Korea would benefit from establishing a more centralized governance structure specifically focused on AI-enabled cyber frauds and modernized regulatory frameworks that address the unique characteristics of generative AI technologies.

National Strategies in Addressing Online Fraud and Scams

There are mainly two types of proactive or preventive policy responses offered by the South Korean government: 1) technical response and reporting system, and 2) public awareness and education.

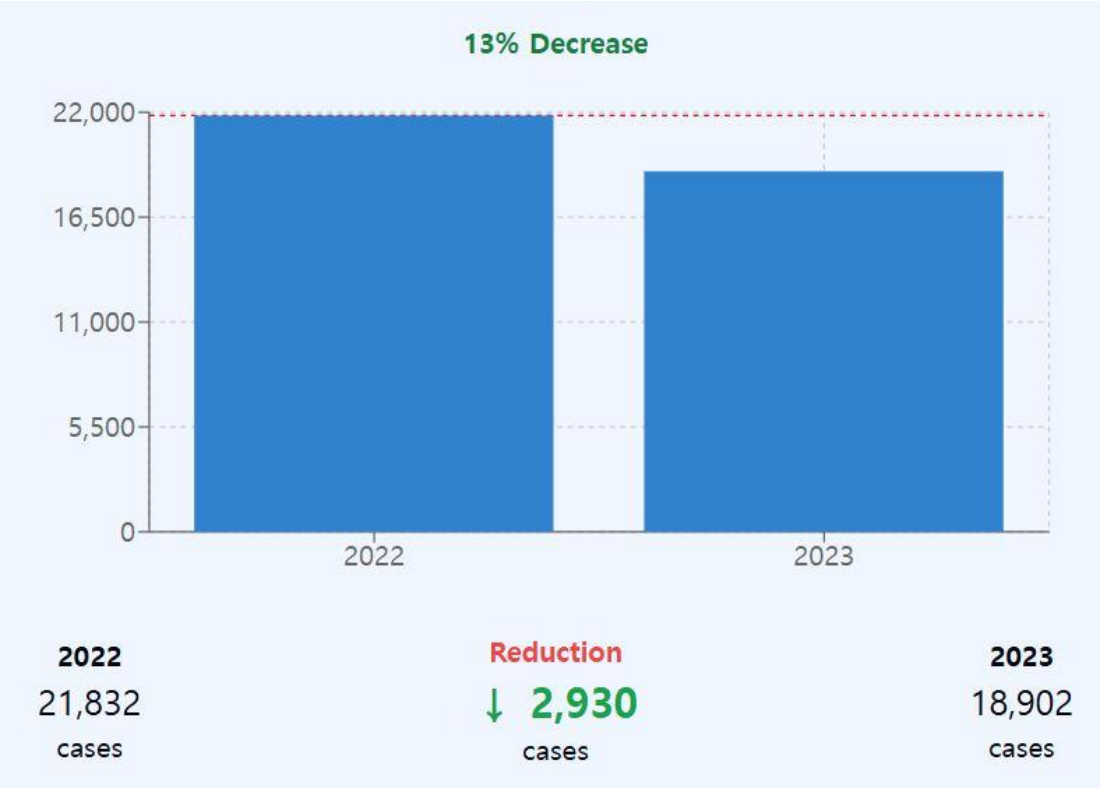
The first approach is technical response; the government developed new technology to detect scam-related activities. For example, the Police University's Public Security Policy Research Institute and the security company InfiniGru jointly developed an app called "Citizen Conan." A Citizen Conan is a mobile security application designed to detect and remove malicious apps that have been installed on smartphones. However, it turned out that criminal actors had already created fake versions of Citizen Conan to target unsuspecting users. This example shows the limitations of technological response alone. Scam enablers and malicious actors continually develop methods to circumvent current anti-scam technology, while users typically bear the burden of exercising caution.

Another example is the Korean National Policy Agency's Electronic Cybercrime Report and Management System (ECRM). Citizens are encouraged to report direct transaction fraud, game transaction fraud, online shopping fraud, and conditional sexual meeting fraud, etc. Currently, the scope of the cybercrime reporting system provided by ECRM extends beyond online scams to include direct attacks on computer and communication networks, infringement of personal and location information, cyber defamation, and

cyberstalking. This suggests that ECRM's efforts may not be sufficient to allocate the necessary resources and attention specifically toward combating online scams, let alone for AI-enabled cyber fraud.

One of Korea's flagship initiatives is the Telecommunications Financial Fraud Integrated Reporting Response Center (TFFIRRC), established in October 2023 under the Korean National Policy Agency. This center serves as a centralized hub for reporting, investigating, and responding to telecommunications-based financial fraud, particularly voice phishing. Korea has implemented systems for analyzing citizen reports and information to identify and block fraud enablers. This approach has yielded measurable results, with voice phishing incidents decreasing by 13% (from 21,832 cases in 2022 to 18,902 cases in 2023) and financial damage reduced by 18% (from 543.8 billion won to 447.2 billion won) (Ministry of the Interior and Safety, March 13, 2024).

Figure 8.6 Voice Phishing Incidents in Korea 2022-2023



Source: Ministry of the Interior and Safety (2024)

While technological and integrated reporting systems form the first line of defense, the Korean government recognizes that an informed public serves as a crucial barrier against cyber fraud. Thus, the second approach focuses on public awareness and education initiatives. Especially during the national holiday season, relevant Ministries and the Ministry of Science and ICT strengthen efforts to combat cyber fraud by publishing Press Releases. Additionally, the Ministry of the Interior and Safety offers safety education videos to prevent fraud and inform citizens on how to report fraudulent activities.

According to the Joint Press Release (2024), for example, the Korean government urged the public to be cautious to minimize damage from various types of cyber fraud ahead of the Chuseok holiday. These include voice phishing, text scams (e.g., smishing), impersonating public institutions for traffic violation fines and illegal dumping penalties, as well as scams impersonating online shopping malls to steal

payments for holiday gifts[21]. This method involves sending text messages containing links to malicious apps, tricking users into installing these apps, or making phone calls that lead to the theft of financial and personal information. Smishing in Korea is often used in crimes, including voice phishing and e-commerce scams.

Other examples include a Card News produced by the Ministry of Science and ICT (MSICT) and other related organizations, with a special campaign during the holiday seasons. For example, the MSCIT provides tips for preventing smishing and online shopping scams during the Lunar New Year holiday season.

Table 8.1 The Korean Government Public Campaigns During Holiday Season

Smishing Prevention Tips	Online Shopping Scam Prevention Tips
• Do not click on unknown URLs or phone numbers. • Enhance smartphone security settings and install apps from official markets. • Install antivirus programs on smartphones. • Never enter or share personal or financial information. • Verify the identity of the requester through phone or video call. • Immediately delete stored photos of ID cards, driver's licenses, and passports from smartphones.	• Verify official shopping malls. • Check business and reviews. • Use secure payment methods. • Check for e-commerce registration.

Source: Joint Press Release (2025). Cyber scam Awareness Campaign

Beyond technical and educational approaches, South Korea has recognized that the borderless nature of cyber fraud, especially AI-enabled scams, necessitates international cooperation. For example, South Korea has expanded its effort to address cross-border frauds and financial crimes systematically. In November 2023, the Korean government hosted the "1st International Conference on Fraud Prevention" in Seoul with participation from 18 countries, including the United States, the United Kingdom, Saudi Arabia, Singapore, and Australia. South Korea also participated in the inaugural "Global Fraud Summit" in London in March 2024 and showed its commitment to international collaboration in fighting transnational fraud. While the summit primarily included G7 and Five Eyes alliance nations, only South Korea and Singapore received special invitations outside these groups (Gov.UK, 2024). At this Summit, 11 major countries adopted the "Global Fraud Summit Communique."

South Korea actively collaborates with Interpol to apprehend fraud suspects and facilitate their repatriation. The government also participates in Interpol funding initiatives for sharing information about criminal organizations involved in telecommunications fraud.

South Korea's multi-faceted approach to combating online scams has demonstrated both promise and persistent challenges. The technical responses, while innovative at first, face continuous evolution of countermeasures by malicious actors, as evidenced by the Citizen Conan case. Educational campaigns raise public awareness but place significant responsibility on individual vigilance. The battle with cyber fraud and AI-enabled scams will ultimately depend not only on reactive measures but on developing

[21] Smishing is a combination of the words SMS (Short Message Service) and phishing.

proactive frameworks that can evolve ahead of emerging threats. The next section explores the potential implications of cyber fraud and online scams.

BEST PRACTICES AND POLICY RECOMMENDATION

The governance limitations become apparent when considering the legal ambiguities surrounding AI-generated harm, particularly in the context of online scams. Various forms of AI-based scams, including deepfake fraud that utilizes celebrities' faces or voices, voice phishing that leverages AI-generated voices, and the dissemination of disinformation, exist in a regulatory vacuum due to ambiguity in determining responsibility. In many specific examples, it is unclear who should be held responsible for the harm caused by AI-generated content. Should the company developing AI technology take responsibility for the results? Or should the platform be responsible for spreading any online scams? Additionally, in cases of cross-border scams, jurisdictional issues persist despite international cooperation. Much of the AI-driven scams are distributed via private or foreign-based platforms (e.g., Telegram), and it is outside the immediate reach of Korean authorities.

The rapid advancement of AI technology has outpaced existing legal frameworks, creating regulatory gray areas in the interpretation, enforcement, and effective response to AI-facilitated financial crimes. When applying existing legal laws to new types of crimes, there are constraints in interpretation and enforcement. Criminal statutes in Korea are being updated to cover AI-facilitated crimes, but practical enforcement —such as tracing sophisticated scams and deepfakes —remains challenging. Even when laws exist, gathering evidence that a video or audio is AI-generated and linking it to a suspect requires advanced forensic expertise.

When it comes to AI-powered voice phishing, for instance, existing laws (such as the Electronic Financial Transactions Act and anti-fraud provisions) mandate that banks implement measures against fraudulent transfers; however, these laws weren't designed with deepfake voices in mind. Verifying a caller's identity is much more challenging when the voice matches perfectly. Under current law, banks are forced to develop new authentication methods.

The challenges of enforcing regulations against AI-driven financial crimes are exemplified by a recent case in Hong Kong, where criminals utilized deepfake technology to impersonate a company executive and authorize a fraudulent transaction worth $25 million. It is unlikely that the employee was terrible at recognizing people's faces. The employee on the call saw what appeared to be their CFO's face and believed the transaction was legitimate. This is new ground for financial oversight bodies and financial institutions. AI technology makes it harder to trace the crime back to its source.

Against this backdrop, South Korea recognizes the importance of collaboration between government agencies and private industry, particularly telecommunications and financial companies, to combat cross-border scams. Most of the time, blocking accounts and communication channels used for fraudulent activities calls for the private sector's willingness to participate. Also, public-private partnerships could be effective for developing joint educational initiatives for consumers.

Forming public-private partnerships for specific domains (e.g. financial fraud, foreign influence operation, deepfake porno) is crucial to combat online scams. The government's primary role is to develop and enforce legal frameworks. With a national legal framework in place, investigating and prosecuting online scammers and criminal networks is essential. Additionally, governments play a key role in promoting and coordinating international cooperation. For private sectors, it could be beneficial to create platform-level fraud detection systems while establishing user protection policies. As AI technology evolves, technology companies are more pressured to develop and implement security technologies. Civil society may support education for vulnerable populations and promote digital literacy.

Sharing best practices would enhance each stakeholder's strategies to react to the AI-generated content with existing fraud schemes. The Korean government, through the Ministry of the Interior and Safety, alongside the National Forensic Service (NFS), has developed an AI-based voice analysis system (K-VoM) to help identify and block voice phishing calls (OPSI, 2024). For this first AI-based voice analysis model in Korea, more than a million Korean and overseas voice datasets from approximately six thousand speakers were utilized (OPSI, 2024). The new model has been applied to the NFS's voice phishing audio analysis process beginning in February 2023, and the police investigation version was distributed to police forces nationwide beginning in July 2023(OPSI, 2024).

AI technology often perpetrates online scams more effectively than it detects and prevents them, leaving stakeholders in a cat-and-mouse game. As AI technology evolves, so do AI fraud schemes, making it harder to react and detect. Technical solutions will inevitably become obsolete over time. While legal frameworks in Korea are beginning to recognize the schemes that AI scammers perpetrate, the regulatory mechanisms to detect and stop these scams in real-time are still catching up.

South Korea already struggles with voice phishing (phone scams) and other cyber-financial fraud, and AI is supercharging these existing schemes in ways that test the defenses of banks and regulators. Imagine a scammer using AI to sound exactly like a family member calling in distress, many people could be convinced to wire money. Banks and consumers can no longer rely on voice recognition or caller ID as proof of identity.

A key enabler is to create an information-sharing system or joint response mechanisms, such as an early warning system that can deploy quickly enough before damage becomes out of control. It is a challenging task to build integrated response systems for large-scale fraud events. This kind of joint response mechanism emerged in South Korea's financial sector.

In February 2025, the Financial Security Institute (FSI) announced proactive measures to enhance the security and reliability of AI applications in the financial sector (2025). This initiative aims to identify security vulnerabilities related to AI technology and improve institutions' fraud detection capabilities. As introduced below, the initiative focuses on creating joint efforts of financial companies against the misuse of AI technology in Korea.

Table 8.2 Financial Security Institute's (FSI) Main Tasks

1. Evaluating security measures of financial companies designated as ***innovative financial services****** for utilizing generative AI, supporting the safe use of AI even in areas where network separation exceptions apply
   *49 services from 32 financial companies has been submitted (as of February 2025)
2. Conducting **AI model security verification** by performing simulated attacks (e.g., using manipulated queries to trick AI into providing incorrect answers or actions) on AI models used by financial companies to identify vulnerabilities, thereby supporting high levels of safety and reliability in AI technology utilization
3. Promoting the development of a **joint AI model for the financial sector** that detects fraudulent financial transactions using the new AI technology of federated learning, expanding the response to fraudulent financial transactions from individual financial company level to a joint financial sector system
4. In addition, planning to support AI utilization in the financial sector in various aspects, including providing an environment where financial companies can easily use **open-source AI models** and supporting the revision of AI guidelines for the financial sector

Building upon South Korea's domestic initiatives, like the FSI's security enhancement program, regional cooperation across Asia presents additional opportunities to strengthen defenses against AI-enabled fraud. Specifically, the integration of AI in fraudulent activities presents unprecedented challenges to the financial security ecosystem and broader society, requiring collaborative responses from the public and private sectors. Asian countries may develop early warning systems to combat international financial scams. Such an initiative would be particularly beneficial in addressing malicious actors located in the Asian region. In addition to early warning systems and information sharing, Asian countries may develop a 'Code of Practice on AI-based Financial Crime', similar to the EU's Code of Practice on Disinformation. This practice could require online platforms to comply with self-regulation to prevent AI-based scams and fraud.

These cooperative regional frameworks offer promising avenues for addressing the governance and regulatory challenges posed by AI-enhanced financial fraud. However, their success will ultimately depend on several critical factors: the speed at which regulatory frameworks can evolve alongside rapidly advancing AI technologies, the willingness of private sector companies to prioritize the cracking down of malicious users over profit, and the development of international enforcement mechanisms.

South Korea's experience with online scams and AI-enabled fraud demonstrates both the evolving nature of digital threats and the challenges of developing effective countermeasures. As AI-enabled cyber frauds infiltrate society, Korean institutions have implemented technical, educational, and regulatory responses with varying degrees of success. Korea's experience highlights the importance of public-private partnerships, international cooperation, and targeted protection for vulnerable populations. Although the nature of crime and the specific technologies used differ, other countries may gain valuable insights by confronting similar challenges - particularly the need for adaptive regulatory frameworks, cross-sector collaboration, and proactive approaches rather than reactive ones.

REFERENCES

Report and Paper

Lesher, M., H. Pawelec and A. Desai (2022), "Disentangling untruths online: Creators, spreaders and how to stop them", OECD Going Digital Toolkit Notes, No.23, OECD Publishing, Paris, https://doi.org/10.1787/84b62df1-en.

National Cyber Security Center, Cyber Threat Analysis Team, "China's Malign Activities by Exploiting "Fake News Websites", NCSC Report-Cyber Threat Analysis (2023)
PdfFileView.do

Korean National Police Agency. 국가수사본부 사이버수사국. Cyber Crime Trend 사이버범죄 트렌드 (2023)

Korea Internet and Security Agency(KISA) 김관영, 김성훈, 이광식, 석지희, 김은성, 이동연. KISA Insight, (2024 Vol. 07, October). <Current Status and Implications of Phishing Response at Home and Abroad: Focusing on the US, EU, UK, Germany, Japan, and China(국내외 피싱(Phishing) 대응 현황 및 시사점: 미국, EU, 영국, 독일, 일본, 중국 중심으로)>

Gov.UK (March 11, 2024) Policy Paper. Global Fraud Summit Communique: 11 March 2024

Article

Jung hun-gu (정헌구). (September 9, 2024). 대한건설경제 지난해 사이버사기 피해액 1조 8,111억원... 4년새 8배 늘었다:대한건설경제

Baek Joon-mu (백준무). (September 9, 2024). Segye Ilbo (세계일보). 사이버사기 피해액, 2023년에만 1조 8000억원... 4년 새 8배 증가

Korean National Policy Agency(경찰청), 대한민국 정책브리핑. "딥페이크 이용한 '자녀 납치' 가짜영상 금융사기 주의" (November 7, 2024)

Oh Hyo-Jung(오효정), The JoongAng. "투자 감사" 조인성 믿었다...수백억 가로챈 가짜 영상의 정체. (February 22, 2024)

YTN, 조인성·송혜교 '투자 권유' 가짜 영상 유포...사기 악용 '논란' [Y녹취록] (February 23, 2024)

Baek Joon-mu & Lee Jian (백준무 & 이지안 기자), Segye Ilbo(세계일보), "○○○인데요" 유명인 사칭 사기 활개... 美선 3000% 급증 [심층기획-사회 혼란 빠뜨리는 '가짜뉴스·딥페이크'] (August 27, 2024)

Observatory of Public Sector Innovation(OPSI) (An official website of the OECD). (July 22, 2024) "Development and Operation of the <Korea Voice Analysis Model(K-VoM) for Voice Phishing>, designed to capture 'Criminal Voices'". Case Study Library

Telecommunications Financial Fraud Integrated Reporting Response Center (counterscam112.go.kr)

Sung hye-mi (성혜미). Yonhap News (연합뉴스) (February 14, 2025). 소비자원 "직구쇼핑몰 사기급증...인스타, 유튜브 연결 67% 달해"

Terrence Matsuo (October 3, 2024) Deepfakes and Korean Society: Navigating Risks and Dilemmas. KEI

https://blog.naver.com/kcc1335/223546792954 SNS로 접근하는 최신 사기 수법! 로맨스 스캠의 예방 및 대처법은? [Source] |작성자 방송통신위원회

Press Release

The Ministry of Science and ICT, "2024년 사이버위협 사례 분석 및 2025년 전망 발표", (December 18, 2024)

Korean National Policy Agency, "2023년 사이버 사기/금융사법 2만 7,264명 검거". (November 23, 2023)

Joint Press Release by Relevant Ministries & Ministry of Science and ICT, (January 19, 2025). Cyber scam Awareness Campaign. "Beware of Cyber Scams Targeting Lunar New Year!"

Relevant Ministries & Ministry of Science and ICT, (January 20, 2025) Card News. "Beware of Cyber Scams Targeting Lunar New Year!"

Financial Services Commission (September 10, 2024). Card News <추석 명절 #보이스피싱 #스미싱 각별히 주의하세요!(예방법, 대응요령)>

Financial Services Commission (January 20, 2025). [보도자료] 설 명절을 겨냥한 문자사기(스미싱) 등 사이버사기 주의

Joint Press Release by Relevant Ministries & Ministry of Science and ICT, "정부, 추석명절 보이스피싱 등 사이버사기 대응 요령 안내" (September 8, 2024)

Joint Press Release by Relevant Ministries & Ministry of Science and ICT, January 19, 2025 보도자료 - 과학기술정보통신부 ("Beware of Cyber Scams Targeting Lunar New Year!"

National Election Commission(중앙선거관리위원회), '딥페이크영상 등' 이용 선거운동 관련 법규운용기준 (2024.1.16)

Ministry of the Interior and Safety(행정안전부). (March 13, 2024), <한국 등 11개 주요국, '초국경 사기범죄방지 성명서' 최초 채택>

Korea Consumer Agency (한국소비자원). (February 14, 2025). <소셜미디어 광고를 통한 해외직구 사기 매년 증가>

Financial Security Institute, '금융보안원, 금융권의 안전한 AI 활용 환경 조성을 위한 보안성 평가 본격 실시' (February 19, 2025)

U.S. Department of Justice's Press Release. (October 18, 2023) Justice Department Announces Court-Authorized Action to Disrupt Illicit Revenue Generation Efforts of Democratic People's Republic of Korea Information Technology Workers

U.S. Department of Justice's Press Release. (January 23, 2025) Two North Korean Nationals and Three Facilitators Indicted for Multi-Year Fraudulent Remote Information Technology Worker Scheme that Generated Revenue for the Democratic People's Republic of Korea

Online Fraud and Scams in Taiwan

Joanna Octavia

9

Online Fraud and Scams in Taiwan

Joanna Octavia[22]

INTRODUCTION

In Taiwan, online scams have risen in prominence, with significant financial losses reported each year. As digital platforms continue to play an increasingly significant role in daily life, scammers have adapted, exploiting the anonymity and vast reach of the internet to target unsuspecting individuals and businesses. Taiwan's susceptibility to online scams is supported by its affluence, an unusually high savings rate of close to 25 per cent, and high internet connectivity, with smartphone penetration rate close to 90 per cent (Fulco, 2025).

Taiwan's scam-related losses continue to escalate, with online investment scams accounting for the largest portion, and social media emerging as the primary platform through which these scams are perpetrated. In 2024, Taiwan experienced a total scam loss of US$ 7.4 billion – the smallest in absolute terms compared to other developed markets in Asia such as Hong Kong and Singapore – but still equivalent to 1 per cent of the GDP (Abraham et al., 2024; Abraham et al., 2024b). Out of this figure, social media postings are becoming the fastest-growing source of scams in Taiwan (Abraham et al., 2024b). More recently, the National Police Agency reported that victims in Taiwan have lost US$46.7 million between March 9 and 15, 2025, with a record of 3790 cases and fraudulent investment schemes accounting for the majority of losses at US$ 25.5 million (Taiwan reports NT$1.46 billion lost, 2025).

To combat this, the Taiwanese government has introduced a range of regulatory measures aimed at curbing the rise of online scams. These include the Fraud Crime Hazard Prevention Act, new frameworks for the virtual asset industry, and version 2.0 of the anti-fraud guidelines. Despite these efforts, challenges remain in addressing the evolving nature of online scams, particularly as scammers continuously adapt their tactics. This case study explores the current landscape of online scams in Taiwan, the regulatory responses put in place to combat them, and recommendations to address the remaining gaps.

CURRENT TRENDS OF ONLINE FRAUD AND SCAMS IN TAIWAN

Taiwan is no stranger to scams. Despite its small population of 23.4 million people, the island is a longstanding hotbed for – and exporter of – telecoms fraud (Chung, 2016). Taiwan is also one of Asia's top manufacturers of high-tech technology, a level of technical expertise that is shared by local scammers (Hale, 2022). While phone calls and text messages remain among Taiwan's top four scam delivery methods, they are now closely followed by social media and instant messaging apps, highlighting a significant shift toward online scams (Abraham et al., 2024b).

In recent years, stronger enforcement against phone-based scams has driven the shift from telecoms fraud to online scams. As traditional telecom frauds faced increased scrutiny and blocking measures, scammers adapted by exploiting social media platforms, online advertisements, and encrypted messaging apps to target victims more effectively. Mirroring global trends, scammers are exploiting artificial intelligence and deepfake technology to make their schemes more convincing and difficult to detect. Coupled with Taiwan's rapid digital adoption during the COVID-19 pandemic, this transition has been further accelerated by regulatory gaps unique to the island, which have made Taiwan a hotspot for sophisticated online scams.

[22] Researcher, Safer Internet Lab, CSIS Indonesia; Associate Lecturer, University College London

Social Media Platforms and Instant Messaging Apps

Social media platforms and messaging apps are central to online scam operations in Taiwan. Scammers use these platforms to build trust and establish relationships before launching scams, often impersonating friends or family members. In addition, phishing and investment scams are also rampant on the platforms. Facebook and LINE are the two most commonly exploited social media platforms, with scammers using them to spread phishing links, impersonate trusted figures, and promote fraudulent investment schemes. In 2023, 55 per cent of Taiwanese respondents reported encountering scam messages through social media, significantly higher than the global average of 44 per cent (Abraham et al., 2024b). In particular, scam activities through social media have increased by 21 per cent between 2023 and 2024 (Abraham et al., 2024b). Despite increasing multi-stakeholder collaborative efforts between government and industry, scams taking place on social media platforms and messaging apps remain difficult to intercept (Hsu et al., 2024).

Facebook stands out as the leading social media platform where individuals encounter online scams. Nearly 70 per cent of online scam losses stemmed from Facebook advertisements, consisting of investment scams and product endorsements by fake celebrities (Yang, 2024). Additionally, 97.9 per cent of fake ads reported to the police were found to have originated from the Meta platform (Yang, 2024). According to the Minister of Digital Affairs, Facebook's centralised and manual approach to removing fake ads has led to a surge in fraudulent advertisements in Taiwan over weekends, when enforcement teams are inactive (Yang, 2024). In November 2024, the prevalence of scams on Facebook led major Taiwanese banks to suspend advertising on the platform in order to safeguard the banks' reputations following a decline in trust toward digital platforms, which is directly attributed to the growing prevalence of scams (Abraham et al., 2024b, Wan, 2024). These developments underscore the importance of more effective measures to tackle the proliferation of fraudulent content on social media platforms.

The widespread use of LINE for scams is a distinctive feature of Taiwan's digital landscape. As Taiwan's dominant messaging app, LINE is used by more than 90 per cent of the population for daily communication, business, banking and shopping (Lange and Lee, 2020). Its popularity extends to official government agencies, banks and brands, many of which operate verified accounts on the platform. Scammers exploit LINE's credibility and trust primarily by creating fake accounts that appear to be from friends, family members, or even trusted organisations. For example, in a typical investment scam, after initial contact on Facebook, victims are directed to fake investment groups on LINE app, where scammers solicit money for fake investment opportunities and asking them to register with fake websites or apps (Su et al., 2024). LINE groups, like other messaging platforms, are private and encrypted, making it challenging for authorities to monitor potential online scam activities once they have entered these private spaces.

Fraudulent investment groups on LINE have been used by scammers to communicate with victims and persuade them to invest in fake investment opportunities. Often branded as stock investment clubs, these fraudulent groups promise exclusive stock tips, cryptocurrency opportunities or insider financial knowledge (Yao and Pan, 2023). A research by Su et al. (2024) comparing fraud and legitimate investment groups on LINE found that the volume of messages in fraud groups exceeded that of the legitimate group, with messages sent predominantly in the afternoon. By creating the appearance of active investment discussions, fraudulent groups sought to influence victims and make them feel more confident about their investment choices (Su et al., 2024). Following closed chat groups, the scammers may continue to engage victims in one-on-one conversations, where they continue to try to persuade victims to invest in fraudulent investment schemes. In response to these tactics, LINE has introduced

warnings across multiple scenarios, such as adding unknown contacts, to alert users to potential scam risks (Ministry of Digital Affairs, 2024).

Phishing attacks attempting to take over LINE accounts for scam activities are common. Users have reported cases where their LINE accounts are being hacked and used to scam their friends and family. A common tactic involves fraudulent invitations to participate in a poll, where victims are prompted to enter their LINE username and password in order to cast their vote (Cheng, 2025; Yao and Yeh, 2025). Once scammers gain access, they use the compromised account to impersonate the victim and send messages to contacts, leveraging personal trust to request money or sensitive information.

Another common scam tactic that exploits the reach of these social networks is by using 'one-page scam posts'. In this tactic, scammers show fraudulent content and links on social media or instant messaging apps, leading victims to other web pages that collect their personal information or trick them into buying fake products (Li, 2024). These one-page scam posts are typically characterised by their use of one long page websites that mix simplified and traditional Chinese characters, and the lack of customer service information (Hiciano, 2025). However, the pages often use names of famous people or organisations, such as the prominent research institution Academia Sinica in Taiwan, to appear legitimate. Moreover, other supporting social media posts that purport similar news are often circulated at the same time to build on the credibility of the fraudulent claims (Li, 2024).

Fake Celebrity Endorsements and Investment Scams

Fake celebrity endorsements are widespread in Taiwan, largely due to the strong idol culture and deep public trust in celebrities. Scammers exploit this trust by using high-profile figures, particularly well-known entrepreneurs, financial experts, politicians and entertainment stars, to lend credibility to their scams and defraud individuals (Yao and Pan, 2023).

One of the most prevalent scam types involving fake celebrity endorsements is investment scams. Alongside fake celebrity endorsements, other common sales tactics used in advertisements for investment scams include free lists of stock picks and high-return investment strategies (Hiciano, 2025). In these scams, users are lured through fake celebrity advertisements on platforms like Facebook and LINE, which feature doctored images or videos of famous individuals endorsing so-called legitimate investment opportunities (Yao and Pan, 2023). Scammers create fake accounts or advertisements featuring their names and images, tricking unsuspecting users into clicking on the links, after which they will try to con victims in the closed chats (Shan, 2023a). In a case involving the impersonation of Lai Xian-zheng, a well-known financial expert, victims were instructed to download a fake app that appeared to be from a trustworthy Japanese securities company and transfer money to a designated bank account (LaMattina, 2024). Many victims believe fake celebrity endorsements are legitimate, as the celebrities used in the ads are widely recognised and respected in Taiwan.

Tackling fake celebrity endorsements in investment scams is crucial, given the significant financial losses associated with these schemes, reputational impact they have on celebrities, and potential scam amplification. In the fourth quarter of last year, investment scams accounted for the largest share of total scam-related losses, at 56.9 per cent of total losses of US$ 1.23 billion (Lee and Pan, 2025). Well-known figures whose images have been used to endorse these scams have filed complaints to the police, leading the police to collaborate with digital platforms such as Google, Meta and others to take down these ads (Yao and Pan, 2024b). The potential for online scam reach and impact to be amplified due to the large followings of these celebrities further underscores the importance of addressing these scams.

Cryptocurrencies

The rise of cryptocurrencies has made it harder to fight online scams in Taiwan. Cryptocurrencies are favoured by scammers targeting Taiwanese users due to reasons such as anonymity, decentralised nature, and ability to operate cross-border (Hung and Van Trieste, 2025). Several unique aspects of Taiwan's cryptocurrency landscape and regulatory environment, such as regulatory gaps in the governance of virtual assets, have made it a significant focal point for crypto-related crime.

Cryptocurrencies are integrated into a typical scam workflow in Taiwan. Since 2019, scammers have started tricking victims by installing fake apps and helping them with cryptocurrency transactions, making the process seem legitimate before stealing their money (Shih and Tsai, 2024). In a typical online scam workflow in Taiwan, online scammers will make initial contact on a public social media platform like Facebook, followed by closed chats as well as one-on-one conversations on LINE. Afterward, scammers will ask the victims to invest in crypto investment scams. After the money is transferred, scammers often attempt to launder it or convert it into cryptocurrency to make it harder for authorities to trace or recover (Lin et al., 2024).

Taiwan's cryptocurrency regulation has historically been relatively light, making it a soft target for crypto-related scams. When markets across Asia such as Singapore, Hong Kong, and China tightened their cryptocurrency regulations, Taiwan positioned itself as a regional crypto hub, attracting investors with its rapidly growing transaction volumes and relatively loose regulatory environment (Tobin, 2023). Taiwan currently does not have a comprehensive regulatory framework for the entire crypto industry, allowing local crypto exchanges to operate with a high level of autonomy. The lack of oversight for crypto exchanges has become a loophole for scammers, while the hands-off stance also meant there was little public education and knowledge on the risks associated with cryptocurrencies and virtual assets.

Misuse of AI in Scams in Taiwan

The rising misuse of AI in online scams is becoming a major concern in Taiwan, especially due to the public's limited awareness of the risks involved. AI-driven deception, which manipulates both visual and auditory elements, introduces new risks for individuals. While many Taiwanese are aware of AI-generated text and chatbots, fewer recognise its use in manipulated images and videos (Abraham et al., 2024b) This is particularly alarming as most scams in Taiwan occur through phone calls and social media, leaving users vulnerable to deepfake images, videos, and voice recordings designed to deceive them.

A significant risk is the use of deepfake technology to impersonate individuals familiar to the victim. This tactic, used by scammers to enhance the credibility of their deception, may build on the widespread phishing of LINE accounts. Furthermore, this tactic aligns with the prevalence of short-term scams in Taiwan, where 39 percent of scams are completed within 24 hours of initial contact (Abraham et al., 2024b). Such concerns over the misuse of AI-altered likeness in scam video calls have been raised by Taiwanese authorities in 2023, following similar trends in China and potential rapid spillover of scams to Taiwan (Liu et al., 2023). This was proven by a case in 2024, whereby scammers used AI to deceive a Taiwanese woman into believing that she was having a video call with Hong Kong celebrity Andy Lau after visiting what she believed was Lau's fan website (Hsu and Pan, 2024).

The usage of AI-generated images and videos to scam individuals in Taiwan is often combined with other technologies, such as social media, online dating sites, or cryptocurrency. This combination has been used in longer-term scams, such as romance and investment scams, which depend on gaining the victim's trust, either through building up a relationship or proving small returns on initial investment. In January 2025, Hong Kong police arrested 31 individuals involved in a syndicate that used deepfake

technology to defraud victims in Taiwan (Ma, 2025). In these romance-investment scams, the scammers created and deployed AI-generated images and videos to impersonate wealthy women and engage victims in online relationships. Using social engineering tactics, the scammers convinced victims to invest in fraudulent schemes, successfully gaining over US$4.3 million as a result (Ma, 2025).

The involvement of technology experts, with their knowledge of AI, machine learning, and data analysis, increases the risk of cross-border scam operations scaling at an alarming rate. The Taiwanese government has raised the concern that online scammers increasingly have the capacity to hire more programmers and obtain greater computing power to run more scams on the internet (Lin et al., 2024). In August 2024, another arrest in Hong Kong led to the capture of a number of digital media and technology university graduates, who were running a global deepfake romance scam operation targeting men around the world, including Taiwan (Yeung, 2024). As scam groups constantly innovate their technologies and tactics, the Taiwanese public, with their limited awareness of AI-generated images and videos in online scams, faces increasing risk of falling for fraudulent schemes.

POLICY GAPS IN ADDRESSING ONLINE FRAUD AND SCAMS

Taiwan has made significant progress in addressing the rising challenges of online scams. In July 2024, the Legislative Yuan passed several key anti-fraud laws, such as the Fraud Crime Hazard Prevention Act, along with amendments to the Code of Criminal Procedure, Communication Security and Surveillance Act, and Money Laundering Control Act. The passage of these laws strengthened regulations on online scams and gave greater power to law enforcement (Chung, 2024a; Executive Yuan, 2024). These were followed by the new version 2.0 of the anti-fraud guidelines (2025-2026), which was passed in November 2024 and primarily focused on financial sector fraud prevention, crypto industry regulations and scam awareness (Executive Yuan, 2024). The new centralised coordination of scam-fighting efforts between agencies, particularly in the monitoring of the financial sector and digital platforms, ensures a more unified and efficient response. However, despite these advancements, there are still policy gaps that need to be addressed for optimal effectiveness.

A big focus of the new policy consists of measures to curb the spread of online scams through online advertising and to increase platform accountability. For example, the real-name system obliges digital platforms to clearly disclose who is running an ad and how the ads are being managed. While this measure could make it harder for scammers to hide behind anonymous ads, it could also expose individual users that are running ads, such as sole proprietors, to privacy risks and safety-related harms. Furthermore, the regulation does not explicitly address user-generated content, such as scam posts made by individual users. This creates a regulatory gap, as online scams can still be perpetrated through non-advertising content.

Meanwhile, the mandatory 24-hour removal of fraudulent ads aims to stop scammers from running harmful ads for long periods and ensures digital platforms take responsibility for removing fraud as soon as they are alerted. This builds on the government's development of an AI-powered system that scans ads and articles on major platforms like Google and Instagram. If the system detects suspicious content, the government will notify the platforms and request prompt removal (Lin et al., 2024). Nonetheless, there may be challenges in ensuring uniform enforcement across all digital platforms, especially when dealing with international platforms outside of Taiwan's jurisdiction. A short turnaround time and insufficient time to thoroughly assess the material can also lead to overblocking of legitimate content.

A key aspect of the new anti-scam guideline is anti-fraud measures in the financial sector, aimed at intercepting fraud via money flows. The FSC is promoting the establishment of mechanisms to detect and flag financial accounts suspected of fraudulent activities, while the mandatory reporting and collaboration among financial institutions facilitates the tracking and freezing of fraudulent accounts

(Executive Yuan, 2024; FSC, 2024). This builds on an existing alliance between the Ministry of Digital Affairs with several partners in the financial industry in Taiwan, including 35 banks (Lin et al., 2024).

Since 2023, the use of deepfakes for online scams has been criminalised in Taiwan. In May 2023, Taiwan's Legislature revised its criminal law to combat deepfake-related scams, raising the maximum prison sentence for those convicted to seven years (Shan, 2023b, Kazaz, 2024). A policy gap here is the limited focus on proactive monitoring and removal of deepfake content that is used to defraud. Furthermore, considering that many online scams involving deepfakes are cross-border with perpetrators based outside of Taiwan, the island's lack of diplomatic ties with some countries, including scam hotspots in Southeast Asia such as Myanmar and Cambodia, make direct prosecution difficult. Additionally, Taiwan is neither a member state nor an observer of Interpol, and therefore is unable to access Interpol's 19 criminal intelligence databases and systems for requesting international cooperation (Coyne, 2024).

The Taiwanese government has moved towards clearer crypto regulations, primarily in response to a significant increase in scams and fraudulent activities within the crypto space, though a fully comprehensive regulation remains to be seen. In 2024, Taiwan introduced new anti-money laundering (AML) for virtual asset service providers (VASPs), mandating them to register by September 2025. Meanwhile, the Financial Supervisory Commission (FSC) is considering new legislation to regulate the cryptocurrency industry (Chung, 2025). Under the new regulations, all VASPs would be classified as financial institutions, and personal trading would be prohibited once the law is enacted (Chung, 2025).

As of March 2025, Taiwan's Financial Supervisory Commission (FSC) has released the draft VASA for public consultation, with feedback open until the end of May. The legislation, expected to be enacted later in 2025, aims to establish a comprehensive regulatory framework for virtual assets. Key provisions include requiring VASPs to join the industry self-regulatory body, which will set and enforce codes of conduct and be subject to oversight by the FSC (Pintu, 2025). However, more details on clear, enforceable consumer protection mechanisms specifically related to crypto-related scams, including the responsibility of VASPs to ensure the security of transactions and prevent scams involving virtual assets, remain to be seen. Furthermore, while the licensing regime ensures that only regulated and compliant VASPs can operate in Taiwan, scammers can still use unregulated foreign crypto exchanges to target Taiwanese users. Tackling this challenge thoroughly will require stronger international cooperation.

POLICY RECOMMENDATIONS

Based on the policy gaps, the following recommendations are proposed:

1. **Expand anti-scam scope to user-generated scam content while ensuring responsible moderation and user protection:** Recognising that online scams in Taiwan increasingly leverage user-generated content (UGC) - such as posts, comments, and fake profiles, the regulatory efforts must broaden their focus beyond just paid advertisements. The government should strengthen mechanisms for timely and fair removal of scam-related UGC through efficient notice-and-takedown procedures grounded in internationally recognised safe harbour principles. It is crucial to adopt a fair approach to content moderation that is transparent and proportionate, protects free speech, and aligns with international internet governance standards. Additionally, this approach should avoid requiring general monitoring of UGC, as it could undermine individual users' privacy and freedom of expression.

2. **Implement privacy safeguards for small advertisers:** To protect privacy, sole proprietors and individuals who run advertisements on digital platforms should be allowed to verify their identity

without publicly disclosing sensitive personal data. One way to do so is by creating anonymised verification badges ("verified advertiser").

3. **Categorise potentially scam ads into tiered risks:** Instead of applying the same 24-hour turnaround time for all flagged content, the government could implement a tiered system where high-risk, clearly fraudulent ads are prioritised for rapid removal. It would require advanced technology, combined with human oversight, to assess the risk levels of flagged content in real time.

4. **Issue proactive regulations to address the misuse of deepfakes:** Taiwan should introduce proactive measures aimed at addressing the misuse of AI and deepfake technologies in scams. These could include mandatory reporting or labelling for deepfake content by digital platforms and promoting AI literacy for consumers to recognise deepfakes.

5. **Create clear guidelines for tackling crypto scams:** The government should introduce more stringent consumer protection regulations for crypto transactions, including ensuring transparency of operations by VASPs and holding them accountable for scam-related activities conducted on their platforms. VASPs should also be required to educate their users on potential risks related to crypto scams.

6. **Strengthen international cooperation:** Given the cross-border nature of online scams, Taiwan should work with international partners to ensure better tracking and removal of deepfake content used in online scams, and to track down and prosecute international crypto scam groups. This could take the form of sharing intelligence and working through international regulatory bodies such as Interpol and other jurisdictions.

REFERENCES

Abraham, J., Rogers, S., Njoki, C., & Greening, J. (2024). *Asia Scam Report 2024.* Global Anti-Scam Alliance. https://www.gasa.org/_files/ugd/7bdaac_e7fad80cd72141c4ac73119be1c9378a.pdf

Abraham, J., Rogers, S., Njoki, C., & Greening, J. (2024b). *The State of Scams in Taiwan 2024.* Global Anti-Scam Alliance. https://www.gasa.org/_files/ugd/7bdaac_479a5775921d4dd598d0af2ca79962e9.pdf

Cheng, J. (2025, January 18). Be more vigilant against scammers. *Taipei Times*. https://www.taipeitimes.com/News/editorials/archives/2025/01/18/2003830404

Chung, J. (2024a, November 29). Cabinet passes new anti-scam guidelines. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2024/11/29/2003827677

Chung, J. (2024b, December 22). Scams cost NT$12.6bn last month: NPA. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2024/12/22/2003828899

Chung, J. (2025, February 2). FSC mulls cryptocurrency regulation. *Taipei Times*. http://taipeitimes.com/News/front/archives/2025/02/02/2003831191

Chung, N. (2016, October 27). Taiwan's cross-strait export of phone scams 'no good for island', former president says. *The South China Morning Post.* https://www.scmp.com/news/china/policies-politics/article/2040413/taiwans-cross-strait-export-phone-scams-no-good-island

Coyne, J. (2024, June 27). *Taiwan's exclusion from Interpol is the world's loss.* Australian Strategic Policy Institute. https://www.aspistrategist.org.au/taiwans-exclusion-from-interpol-is-the-worlds-loss/

Executive Yuan. (2024, December 6). *Next-generation anti-fraud strategy guidelines, version 2.0* [Press release]. https://english.ey.gov.tw/News3/9E5540D592A5FECD/faccc48c-1d4c-45c8-aa1b-73d9c283a73d

Financial Supervisory Commission. (2025, January 22). *Measures to enhance fraud prevention in Taiwan's financial sector* [Press release]. https://www.banking.gov.tw/en/home.jsp?id=87&parentpath=0&mcustomize=multimessage_view.jsp&dataserno=202501220001&dtable=News#:~:text=In%20terms%20of%20fraud%20prevention,accounts%2C%20(3)%20Requiring%20card

Fulco, M. (2025, February 17). *Taiwan grapples with intensifying cybercrime.* Taiwan Business TOPICS. https://topics.amcham.com.tw/2025/02/taiwan-grapples-with-intensifying-cybercrime/

Hale, E. (2022, December 5). Taiwan's front-line battle against mobile phone fraud. *BBC.* https://www.bbc.co.uk/news/business-63075729

Hiciano, L. (2025, February 19). MODA shares scam warning signs. *Taipei Times.* https://www.taipeitimes.com/News/taiwan/archives/2025/02/19/2003832150

Hiciano, L. (2025, February 20). MODA reveals common online scam key words. *Taipei Times.* https://www.taipeitimes.com/News/taiwan/archives/2025/02/20/2003832194

Hsu, S. & Pan, J. (2024, July 2). Scammers use AI to cheat woman out of NT$2.64m. *Taipei Times.* https://www.taipeitimes.com/News/taiwan/archives/2024/07/02/2003820211

Hsu, H-C., Gillespie-Jones, C., Kaushal, A., & Yasukagawa, K. (2024, August 21-23). *Messaging Scam and Combatting to Protect Human Rights and Democracy* [Conference presentation]. Asia-Pacific Regional Internet Governance Forum, Taipei, Taiwan. https://aprigf.tw/programs/messaging-scam-and-combatting-to-protect-human-rights-and-democracy/

Kazaz, J. (2024). *Regulating Deepfakes: Global Approaches to Combating AI-Driven Manipulation.* GLOBSEC. https://www.globsec.org/what-we-do/publications/regulating-deepfakes-global-approaches-combatting-ai-driven-manipulation

Lange, E. & Lee, D. (2020, November 23). How One Social Media App is Beating Disinformation. *Foreign Policy*. https://foreignpolicy.com/2020/11/23/line-taiwan-disinformation-social-media-public-private-united-states/

LaMattina, L. (2024, August 8). Police investigate 27 suspects over NT$400 million investment fraud in north Taiwan. *Taiwan News.* https://taiwannews.com.tw/news/5916118

Lee, W. & Pan, J. (2025, January 20). Investment scams topped losses from fraud in 4th quarter. *Taipei Times*. https://www.taipeitimes.com/News/front/archives/2025/01/20/2003830510

Li, W. (2024, April 29). Too good to be true - Online scams in Taiwan. *Taiwan FactCheck Center*. https://en.tfc-taiwan.org.tw/too-good-to-be-true-online-scams-in-taiwan/

Lin, Y., Sahel, J., Chan, J., Octavia, J., & Chung, E. (2024, August 21-23). *From Innovation to Impact: Responsible AI - Challenges and Opportunities to Tackle Online Fraud and Scams* [Conference presentation]. Asia-Pacific Regional Internet Governance Forum, Taipei, Taiwan.

Liu, C., Lin, C., & Madjar, K. (2023, May 30). Police warn against scams using deepfake videos. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2023/05/30/2003800665

Losses from scams hit NT$239.5bn. (2024, October 2). *Taipei Times.* https://www.taipeitimes.com/News/front/archives/2024/10/02/2003824665

Ma, J. (2025, January 5). Hong Kong police arrest 31 over deepfakes used to scam victims in Singapore, Malaysia. *The Straits Times*. https://www.scmp.com/news/hong-kong/law-and-crime/article/3293476/hong-kong-police-arrest-31-who-used-deepfakes-scam-victims-singapore-malaysia

Ministry of Digital Affairs. (2024, June 26). *Ministry of Digital Affairs Coordinate with Google and LINE to Strengthen Online Fraud Prevention Measures* [Press release]. https://moda.gov.tw/en/press/press-releases/13010

Pintu. (2025, March 27). *New Regulation of Crypto Assets in Taiwan: What Impact for Investors?* https://pintu.co.id/en/news/143170-new-regulation-of-crypto-assets-in-taiwan

Shan, S. (2023a, May 2). Google joins effort to combat online scam ads in Taiwan. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2023/05/02/2003798981

Shan, S. (2023b, May 17). Legislature passes stiffer jail, fine for deepfake fraud. *Taipei Times*. https://www.taipeitimes.com/News/front/archives/2023/05/17/2003799936

Shen, T. (2024, November 28). Taiwan fast-tracks stricter crypto AML rules to take effect Nov. 30. *The Block*. https://www.theblock.co/post/328729/taiwan-fast-tracks-aml-rules-stricter-crypto

Shih, C. & Tsai, M. (2024). Application of Money Flow Analysis Technology in the Investigation of Money Laundering Crimes in Taiwan. *Procedia Computer Science*, 246, 4524-4533.

Su, Y., Shih, C., & Yang, T. O. (2024). Investment Fraud Cases Study in Chinese Context of Instant Messaging Software. *28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024).* Procedia Computer Science 246 (2024) 391-402.

Taiwan reports NT$ 1.46 billion lost to scams in one week. (2025, March 27). *TVBS News.* https://news.tvbs.com.tw/english/2810509?from=english_content_pack

Tobin, M. (2023, January 3). Taiwan goes all in on crypto, despite the global crash. *Rest of World*. https://restofworld.org/2023/taiwan-asia-crypto-capital/

Wan, C. (2024, November 19). Taiwanese Banks Curtail Facebook Ads Over Scam Concerns. https://www.bloomberg.com/news/articles/2024-11-19/taiwanese-banks-curtail-facebook-advertising-over-scam-concerns

Yang, J. (2024, December 25). Facebook at the Heart of Why Taiwan Can't Stop Scams. *Commonwealth Magazine*. https://english.cw.com.tw/article/article.action?id=3889

Yao, Y. & Pan, J. (2023, December 19). Stock investment scams are on the rise, bureau warns. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2023/12/19/2003810846

Yao, Y. & Pan, J. (2024, February 24). Most online fraud on Facebook: report. https://www.taipeitimes.com/News/taiwan/archives/2024/02/24/2003814012

Yao, Y. & Pan, J. (2024, August 6). Meta pulls more than 46,500 fraud ads. https://www.taipeitimes.com/News/taiwan/archives/2024/08/06/2003821850

Yao, Y. & Yeh, E. (2025, March 19). CIB issues warning about hijacked social media accounts. *Taipei Times*. https://www.taipeitimes.com/News/taiwan/archives/2025/03/19/2003833689

Yeung, J. (2024, October 15). Deepfake scams raked in $46 million from men across Asia, police say. *CNN.* https://edition.cnn.com/2024/10/15/asia/hong-kong-deepfake-romance-scam-intl-hnk

Online Fraud and Scams in Thailand

Saliltorn Thongmeensuk

10

Online Fraud and Scams in Thailand

Saliltorn Thongmeensuk[23]

PATTERNS AND TRENDS

Online fraud and scams have become a significant issue in Thailand's digital landscape. According to the Cyber Crime Investigation Bureau, approximately 700 cases of online fraud and scams are reported daily (Siam Legal, 2025). These crimes are carried out using various methods. Statistics from the Anti-Online Scam Operation Center (AOC 1441) reveal that between 2024 and 2025, the five major online fraud cases resulted in a combined total loss of $610,000. Scammers employed different tactics to deceive victims, including fraudulent social media posts or messages, misleading online platform advertisements, fake job offers, and deceptive phone messages. These approaches exploit social media platforms and digital communication channels to manipulate individuals into financial losses, highlighting the urgent need for stricter online security measures and public awareness campaigns.

Moreover, data from the Global Anti-Scam Alliance's survey indicates that in 2024, scammers primarily used scam calls, text messages, social media, and online advertisements to target victims. Additionally, with the rise of AI-powered text generation, a majority of survey participants reported that more than half of the scam messages they received appeared to be AI-generated, making them more sophisticated and difficult to detect. The survey also revealed an increasing trend in identity theft, where scammers impersonate government officials or law enforcement officers to manipulate victims into providing sensitive information or making fraudulent payments (Nation Thailand, 2024).

The growing trend of identity theft has become even more alarming with the emergence of AI-assisted identity fraud, which significantly enhances the sophistication of these crimes. In February 2025, a notable case of AI-assisted identity theft occurred, where scammers used AI-powered face-altering technology during video calls to make their deception more convincing. This scheme successfully tricked 163 victims, including a well-known influencer who suffered a financial loss of over four million baht (Bangkok Post, 2025a). Another shocking incident took place in January 2025, involving an AI-assisted scam call targeting Paetongtarn Shinawatra, the current Prime Minister of Thailand. In this case, scammers attempted to extort money by impersonating a well-known world leader using AI-generated voice cloning technology. However, Paetongtarn became suspicious of the voice's authenticity and did not fall victim to the scam. Despite avoiding the trap, the prime minister later stated that the AI-generated voice was highly convincing, highlighting the growing threat posed by advanced deepfake technology in cybercrime (Nation Thailand, 2025a).

Another infamous case reported by the researchers from cyber security firm – Group-IB - is when the group of hackers -GoldFactory- try to use facial recognition AI to steal money from the victim. In 2023, the group of hackers upload the application called "Digital Pension" masked as legitimate service. The application requires the victims to record the video to access the pension, then the fraudster will use facial recognition AI to record the face of the victim and use face swap AI solution to create a deepfake video (Group-IB, 2024). This pattern emerged when Bank of Thailand mandated that any digital money transfer exceeding 50,000 baht per transaction, or daily transfers surpassing 200,000 baht, require biometric authentication, specifically facial recognition, to verify the account owner's identity. This policy aimed to bolster security for high-value transactions to combat online fraud and scam (Bank of Thailand, 2023a), but with exploitation of AI technology the fraudster can easily overcome this measure.

[23] Senior Research Fellow, Thailand Development Research Institute (TDRI)

These incidents raise significant concerns about the increasing volume and sophistication of online fraud and scams, as artificial intelligence is being exploited to enhance deception. With AI-powered scams becoming more convincing, fraudsters can now manipulate voices, videos, and messages with alarming accuracy, making it increasingly difficult for victims to distinguish between legitimate and fraudulent interactions.

The rapid evolution of AI-driven cybercrime underscores the urgent need for government intervention. Authorities must implement stricter regulations, enhance cybersecurity measures, and invest in AI-driven fraud detection technologies to combat this growing threat. Additionally, public awareness campaigns, digital literacy programs, and cross-border cooperation are essential to protect individuals from falling victim to these increasingly sophisticated scams. Without proactive action, the risk of financial and personal losses due to AI-assisted fraud will continue to escalate.

THE ROLES OF KEY STAKEHOLDERS IN ADDRESSING ONLINE FRAUD AND SCAMS

The issue of online fraud and scams in Thailand involves multiple stakeholders, each playing a critical role in prevention, enforcement, and victim assistance. However, while existing initiatives have proven effective in mitigating fraud, challenges remain in proactively addressing emerging threats, particularly AI-enabled scams.

Ministry of Digital Economy and Society (MDES)

As the primary coordinating body in the fight against online fraud, the Ministry of Digital Economy and Society (MDES) has taken a central role in policy formulation, enforcement coordination, and technological innovation. MDES has established collaborations with multiple agencies, including:

- **The Royal Thai Police's Cyber Crime Investigation Bureau** – responsible for investigating and prosecuting cybercriminals.
- **The Anti-Money Laundering Office (AMLO)** – tracks illicit financial flows linked to scam networks.
- **The Bank of Thailand (BoT) and the Thai Bankers Association** – oversee fraud prevention measures within the banking sector.
- **The Department of Special Investigation (DSI) and the Securities and Exchange Commission (SEC)** – target financial fraud, investment scams, and cyber-enabled crime.
- **The National Broadcasting and Telecommunications Commission (NBTC)** – regulates telecommunication networks, focusing on scam calls, SMS fraud, and telecom-related cybercrime.

Recognizing the need for a centralized and rapid response mechanism, MDES, in partnership with these organizations, established the Anti-Online Scam Operation Center (AOC 1441). The AOC serves as a one-stop service to combat online scams, allowing victims to report fraudulent activities and enabling immediate intervention through a 24/7 hotline (1441). While AOC 1441 has significantly improved victim support and scam mitigation, it remains a reactive measure—primarily remedying fraud after it occurs rather than preventing scams in advance. This limitation has led MDES to develop more proactive fraud detection solutions.

To enhance prevention-oriented approach, MDES is developing the DE-fence platform, set to launch in late 2025. This initiative aims to detect fraudulent patterns before scams occur, using artificial intelligence (AI) and big data analytics to monitor and analyze suspicious financial activities, scam websites, and digital transaction behaviors. The DE-fence system will integrate with AOC 1441, expanding its role from reactive intervention to real-time scam detection and prevention. Key expected capabilities include:

- Automated fraud detection – leveraging AI to flag scam patterns in real-time.
- Telecom and banking fraud surveillance – identifying suspicious SIM registrations and fraudulent transactions.
- Predictive threat modeling – analyzing emerging scam tactics to inform regulatory responses.

The Bank of Thailand (BoT) and Its Role in Fraud Prevention

As the primary financial regulator, the Bank of Thailand (BoT) has been instrumental in mandating security policies for commercial banks to minimize fraud risks. Given that the banking sector accounts for a majority of online financial fraud cases, BoT has introduced several countermeasures, including:

1. Mandatory Multi-Factor Authentication for High-Value Transactions
    - In mid-2023, the BoT has required facial recognition or fingerprint authentication for transactions exceeding 50,000 THB per transfer or 200,000 THB per day.The measure was introduced to prevent unauthorized transactions and account takeovers.

2. Continuous Adaptation to Emerging Fraud Patterns
    - BoT acknowledges that online financial fraud evolves rapidly—when regulations are enforced, scam activities decrease temporarily before criminals adapt with new tactics. To counter this, BoT has actively updated banking security frameworks, including guidelines for commercial banks on digital fraud prevention and investment in AI-driven security tools. In addition, BoT has encouraged financial institutions to strengthen biometric verification systems against AI-based identity fraud.

3. Cross-Sector Collaboration to Combat Scams
    - BoT also works closely with MDES, law enforcement agencies, and commercial banks to formulate fraud mitigation strategies.It has also emphasized cross-border intelligence sharing to tackle transnational fraud networks.

Thai Bankers Association (TBA) and The Central Fraud Registry (CFR)

Recognizing the importance of tracking financial crime across multiple institutions, the Thai Bankers Association (TBA) launched the Central Fraud Registry (CFR) in mid-2024. This initiative aims to track and eliminate mule bank accounts used for money laundering and scam operations.

Commercial Banks and E-Payment Platforms

Commercial banks and e-payment service providers play a critical role in fraud prevention, transaction monitoring, and risk assessment. Under BoT's directives, financial institutions are:

- Required to implement transaction monitoring systems capable of detecting and freezing suspicious activities in real-time.
- Obliged to freeze accounts linked to fraudulent transactions and report them to the Central Fraud Registry.

- Encouraged to invest in AI-driven fraud detection to combat deepfake-assisted identity theft and AI-powered phishing scams.

Telecommunications Sector and The Role of NBTC

Given that many online fraud schemes originate from scam calls and fraudulent SMS messages, the National Broadcasting and Telecommunications Commission (NBTC) has implemented:

- Restrictions on bulk SIM card registrations to prevent scammers from using untraceable phone numbers.
- Telecom providers are obliged to block known scam numbers and filter fraudulent messages before reaching consumers.
- Collaboration with law enforcement to track down illegal call centers, many of which operate from neighboring countries.

POLICY OVERVIEW IN ADDRESSING ONLINE FRAUD AND SCAMS IN THAILAND

Anti-Online Scam Operation Center

The most recent government policy to combat online fraud and scams was introduced in 2023, when the Ministry of Digital Economy and Society (MDES), in collaboration with various public and private sector organizations, established the Anti-Online Scam Operation Center (AOC 1441). This initiative serves as a one-stop service designed to combat online scams using an AI-assisted platform capable of analyzing scam patterns, identifying fraudsters, and swiftly taking down their bank accounts (Ministry of Digital Economy and Society, 2023).

The online fraud and scam often involve several agencies to tackle the case effectively, the victim will need to contact several agencies, fill several documents and wait for month to conclude the case. However, the AOC was established from the collaboration of all the responsible agency, namely MDES (which oversees the center), the Royal Thai Police's cybercrime units, the Anti-Money Laundering Office (AMLO), the Bank of Thailand and Thai Bankers Association, the Department of Special Investigation (DSI), the Securities and Exchange Commission (SEC), and the National Broadcasting and Telecommunications Commission (NBTC) (Nation Thailand, 2024b), thus considerably shorten the time needed to tackle the case from days to withing an hour.

The AOC operates a dedicated hotline service with 100 active lines, ensuring that victims can report scams 24/7. The center has set an ambitious response time, needing only 10 minutes to investigate cases, freeze fraudulent accounts, and facilitate the return of stolen funds to victims (Nation Thailand, 2024b). Furthermore, AOC is employed AI and big data to combat fraudulent activities swiftly, the AI will analyze massive data stream from several sources – most of which is provided by the center's partner e.g., Central Fraud Registry of the Thai Bankers' Association and telecom data exchange system (Bangkok Post, 2024a). This technology allows AOC to take down the scammers bank account swiftly, and ensure the victims will get their money back within an hour after the call (Bangkok Post, 2024a).

Since its launch in November 2022, the AOC has demonstrated significant effectiveness in fighting cybercrime. By February 2025, the center reported receiving over one million calls from scam victims and successfully taking down 537,431 scammer-linked bank accounts. This rapid intervention has dramatically curbed the impact of online fraud. Officials reported that the daily financial damage from scams fell by 36–44% after the 1441 hotline became active (Bangkok Post, 2025b). This achievement highlights the Thai government's proactive efforts in leveraging technology, cross-sector collaboration,

and rapid-response mechanisms to mitigate the growing threat of online fraud (Royal Thail Government, 2025).

Cybersecurity Measures

Apart from AOC, Thailand has enacted the Royal Decree on Measures for Protection and Suppression of Technology Crimes B.E. 2566 (2023) ("Cybercrime Decree") which give the rights to the victims and authorities to ask the commercial bank and e-payment service to freeze suspicious crimes for seventy-two hours, then within the seventy-two hours the victim must file the complaint to the police. Then the police need to notify the bank or platform of the complaint, and the bank or platform need to freeze the account for another seven days, the account can be unfrozen after seven days if no further extension order from the police. Furthermore, the bank or platform needs to proactively freeze the suspicious activities they find for seven days and notify the authorities, it can be unfrozen after seven they if there is no extension order (Tilleke & Gibbins, 2025).

Moreover, Thai cabinet approved the draft amendment to the Executive Decree on Measures to Prevent and Suppress Technological Crime 2023 which will add several obligations to the P2P cryptocurrency trading platform, commercial bank, telecom operators and social media platform to exercise due care to prevent online fraud and scam (Nation Thailand, 2025b). This amendment is expected to take effect in early 2025.

Another key figure to combat online fraud and scams are the Bank of Thailand (BOT) the regulator in banking sector and central bank of Thailand. The BOT took several proactive measures to combat online scam e.g., the central bank mandated the rule for every private bank to set up hot line to tackle online scam. Moreover, the BOT mandated a rule in 2023 to prevent the fraudulent activities in which the scammer steal one-time-password (OTP) or log-in details to siphon money from the victims (Bank of Thailand, 2023b).

The Use of AI in Addressing Scams

While AI technology is one of the key success factors to combat and effectively reduce online fraudulent activities and redistribute the stolen fund back to the victim, the exploitation of AI technology complicated the pattern on online fraud and scam. From the case mentioned in the previous section, the GoldFactory – a Chinese Hacker group – exploited the BOT's biometric prevention policy using facial recognition technology and deepfake to impersonate the victim and siphon the money out of the bank. This raises concern especially for the Bank of Thailand to tackle the use of Deepfake, the Bank of Thailand adds two more layers requires for the transaction, to make a large transaction one will need to verify with his phone, 6-digit OTP and face scan before concluding the activity. The fraudster will need to obtain the victims phone number, intercept the message to get the OTP, then use face scan to conclude the transaction, which makes the siphoning activities harder. Moreover, BOT and Thai commercial Banks are working to enhance its facial recognition technology to detect deepfake video more accurately (Bank of Thailand, 2025).

Furthermore, Thai commercial banks have been proactively investing in technology to detect fraudulent activities and works together to create a database called "Central Fraud Registry" which allows banks to track fraudulent activities across banking sector. As a result, the banking sector in Thailand are developing quickly to prevent online fraud and scam online.

In telecommunication side, National Broadcasting and Telecommunication Commission mandated a rule that any user holding more than 5 sim card needs to verify his usage or the sim cards will be deactivated. This policy aims to tackle the call center gang who uses scam call to trick the victims.

Regional Collaboration in Addressing Scams

The nature of the online fraud and scam often involve the crime conducted by the fraudster outside the country, which requires the cooperation from several countries to tackle the problem effectively. In 2025, ASEAN Digital Ministers' Meeting in Bangkok, Paetongtarn Shinawatra the prime minister of Thailand states that prevention of online scam must be one of the priorities for ASEAN countries, and the use of Artificial Intelligence need to follow an ethical guideline. Consequently, the meeting concluded with the decision to strengthen the collaboration among ASEAN countries to combat online crime through ASEAN Working Group on Anti-online Scam (WG-AS) and strengthen AI governance through ASEAN Working Group on Artificial Intelligence Governance (WG-AI) (ASEAN, 2025). Moreover, the ASEAN countries will employ ASEAN Cybersecurity Coordinating Centre to share data regarding the pattern of AI-enabled scam among members to introduce new measure to combat the problem (ASEAN, 2025).

Moreover, Thailand and China work together to tackle the scammers located in neighboring countries like Myanmar, Cambodia, and Laos. In a high-level agreement in January 2025, Thailand and China announced they will set up a joint coordination center to combat illegal call center networks along the Thailand-Myanmar and Thailand-Cambodia borders. The coordination center is based on Bangkok and the other coordination center will be established in Mae Sot. The center brings Thai and Chinese police together to share intelligence and co-direct operations against the syndicates (Reuters, 2025). The joint effort is successful with a coordinated crackdown in early 2025 led to the rescue or detention of over 7,000 people from scam centers in Myanmar's Myawaddy region, many of them trafficked workers forced to run scams (The Irish News, 2025).

IMPLICATIONS OF ONLINE FRAUD AND SCAMS

Prior to the implementation of targeted interventions, online fraud and scams in Thailand had escalated to unprecedented levels. According to the Cyber Crime Investigation Bureau, from March 2022 to May 2024, Thailand recorded approximately 300,000 cases of online financial fraud, resulting in total economic losses exceeding 63 billion THB (approximately USD 1.8 billion) (Royal Thai Police, 2023).

Beyond direct financial losses, fraudulent activities have had a broader economic impact. Online scams led to reduced consumer confidence in digital financial services, directly affecting e-commerce growth and financial technology (fintech) adoption. A survey conducted by the Electronic Transactions Development Agency (ETDA) in 2023 revealed that 43% of Thai respondents expressed reluctance to conduct online transactions due to fraud concerns, while 29% indicated that they had reduced their usage of mobile banking services following exposure to scam attempts (ETDA, 2023).

In response to that, Thai authorities implemented a series of countermeasures between 2022 and 2024, focusing on financial security, law enforcement coordination, and cross-border cooperation. To enhance security in digital transactions, the Bank of Thailand (BoT) introduced mandatory biometric verification for transactions exceeding 50,000 THB per transfer (or 200,000 THB per day) in mid-2023 By requiring facial recognition or fingerprint authentication, this policy aimed to reduce unauthorized access to digital banking services. The outcome of this intervention has yet to be publicly announced.

Moreover, The Central Fraud Registry (CFR) was introduced in 2024 to consolidate fraudulent account data across banks. By mid-2024, Thai banks collectively closed 1.8–1.9 million mule accounts (ETDA, 2023), reducing the ability of scammers to launder stolen funds. Though the intervention has taken down several mule accounts, but those accounts increase much more substantially and still experience an increasing trend which call for stronger intervention.

Furthermore, Bank of Thailand's report shows that the online financial fraud evolves quickly, when there is a new intervention, the fraud will reduce significantly until the new generation of online fraud emerge calling for the authorities to introduce a new regulation, which will eventually reduce the new generation fraud pattern. This trend shows that the regulation regarding online fraud and scam needs to be constantly updated to keep up with the new pattern of fraudulent activities (Bank of Thailand, 2024).

Apart from the effort exclusively in the banking sector, the joint effort among stakeholders involving online fraud and scam has been introduced as the establishment of The Anti-Online Scam Operation Center (AOC 1441) was established in late 2023 to enable rapid intervention in fraud cases. The center's 24/7 hotline allowed victims to report scams in real-time, triggering an immediate freeze of suspect accounts within an average response time of 10 minutes (Edulampang, 2024). The impact of this intervention recorded from November 2023 to February 2025, the AOC received 1,461,074 calls and froze 517,954 accounts (Edulampang, 2024).

However, these interventions aim to tackle online fraud and scam activity in more traditional ways which involve text message, scam call, online advertising and social media platforms. Meanwhile, advanced technology like Artificial Intelligence will make fraud and scam become much more sophisticated and harder to detect.

The increase in exploitation of AI technology in organized crime might be the result of the increase in accessibility of the technology. The emergence of AI and Deepfake technology lead to an increase in open-source AI tools and acceleration in Generative AI deployments make AI technology more accessible, which leads to emerging risks from AI technology as follows (UNODC, 2024):

- Deployment in cyber-enabled fraud: advanced technology provide the scammer with more way to conduct the fraudulent activities e.g., automatic scripts for social engineering, impersonating scam, KYC bypass.
- Evolution and Scalability: the rapid deployment of Generative AI enable less tech-savvy scammers to conduct cyber-enabled fraud, make the cyber-enable fraud becomes much more scalable (automatically generate text messages).

The APAC region experiences a rise in cyber-fraudulent involving artificial intelligence such as the rise in Deepfake-related crime in 2023 which increase by more than 1500 percent. Thailand also experienced the rise in these crimes, which include:

- Deepfake fraud (impersonation of CEOs, government officials, or family members)
- AI-generated phishing emails and voice cloning scams
- Face-swapping technology to bypass biometric authentication

In a recent case, deepfake scam where fraudsters used AI to impersonate police officers during a video call which results in more than 190 victims, Moreover, one of the most shocking cases is the voice cloning scammers impersonate the world leader's voice and try to extort the prime minister of Thailand. Another case is the attempts of group of hackers trying to steal biometric data from Thai citizen involving launching an application luring the victims to upload their video and use facial recognition AI to capture the biometric data of the victim to siphon money from their account. For these reasons, though there might not be the statistics pointing out the exact impact of AI-enabled fraud in Thailand, but the cases from these fraudulent activities raises a concerning for the use of AI in online fraud and scam.

BEST PRACTICE AND POLICY RECOMMENDATIONS

Thailand initially faced high volumes of online fraud, but through multi-stakeholder collaboration, led by the Ministry of Digital Economy and Society (MDES), the country has successfully mitigated these threats. The establishment of AOC 1441 was a major step in reducing fraud, enabling faster intervention and enhanced coordination between government agencies. The Bank of Thailand's biometric authentication mandate has shown promise in preventing traditional fraud. However, scammers quickly adapt to new regulations, requiring continuous policy adjustments to stay ahead of evolving threats.

In terms of public-private collaborations, Google and the Ministry of Digital Economy and Society (MDES) partnered to enhance Google Play Protect in April 2024 (Bangkok Post, 2024b). The partnership has resulted in the blockage of more than 6.6 million high-risk app installation attempts as of April 2nd, 2025. Additionally, Thailand's National Cyber Security Agency (NCSA) and Google Cloud have also announced a strategic collaboration and engage in AI-powered cyber defense through threat intelligence sharing and incident response capability building to address evolving threats and boost online safety for Thailand citizens and residents (Google, 2025).

At the international level, the joint Thai-Chinese crackdown on cybercrime near the Myanmar border was a notable success, highlighting the importance of cross-border cooperation. Additionally, the establishment of ASEAN's working group on AI and online fraud prevention is a significant step toward ensuring regional cybersecurity.

Despite these efforts, AI-powered fraud remains an emerging threat in Thailand. While the country has introduced AI ethics guidelines, they lack regulatory enforcement mechanisms, leaving high-risk AI applications unregulated. There are several key actions that need to be taken in addressing online fraud and scams in Thailand:

- A combined effort among government agencies within the country is crucial to tackling online fraud and scams. This approach requires a robust database and advanced analytical tools (e.g., artificial intelligence) to identify crime patterns and enable swift action.
- Prevention-oriented measures are necessary to enhance online safety, but such an approach requires cutting-edge technology and data analytics for effective implementation.
- International collaboration is essential, as online fraud often involves cross-border organized crime networks.
- Online fraud evolves rapidly with the increasing accessibility of generative AI and advanced AI solutions. Stakeholders must continuously adapt to counter emerging threats. Consequently, data sharing and best practices between countries are critical to ensuring well-informed, timely interventions.
- Proactive partnerships with key actors in the private sector (e.g., banks, digital payment providers, telecommunication companies, e-commerce platforms, social media) should be leveraged to enhance industry expertise and strengthen joint responses for early detection, mitigation, and prompt scam responses.

REFERENCES

ASEAN. (2025). *Bangkok Digital Declaration.* ASEAN. Retrieved from https://asean.org/wp-content/uploads/2025/01/14-ENDORSED-BANGKOK-DIGITAL-DECLARATION.pdf.

Bangkok Post. (2024a, December). Govt sets up scam victim aid. Retrieved from https://www.bangkokpost.com/thailand/general/2676383/govt-sets-up-scam-victim-aid

Bangkok Post. (2024b, December). State-Google fraud effort blocks scams. Retrieved from https://www.bangkokpost.com/business/general/2915570/state-google-fraud-effort-blocks-scams

Bangkok Post. (2025a, February). Two men arrested for alleged B4m AI-aided scam against beauty queen. Retrieved from https://www.bangkokpost.com/thailand/general/2953450/two-men-arrested-for-alleged-b4m-ai-aided-scam-against-beauty-queen.

Bangkok Post. (2025b). Digital Economy and Society to retain key policies. Retrieved from https://www.bangkokpost.com/business/general/2866637/digital-economy-and-society-ministry-to-retain-key-policies

Bank of Thailand. (2023a, March). Bank of Thailand implements new cybersecurity measures. Retrieved from https://www.bot.or.th/en/news-and-media/news/news-20230309.html

Bank of Thailand. (2023b). Regulatory update: Strengthening financial cybersecurity. Retrieved from https://www.bot.or.th/en/news-and-media/news/news-20230309.html

Bank of Thailand. (2024). *Improving measures to manage financial fraud.* Bank of Thailand. Retrieved from https://www.bot.or.th/content/dam/bot/documents/th/news-and-media/news/2024/news-th-20240613-attach1.pdf

Bank of Thailand. (2025, February). Many scammers are closely following the new scams with better protection methods. Retrieved from https://www.bot.or.th/th/research-and-publications/articles-and-publications/bot-magazine/Phrasiam-68-1/TheKnowledge_cofact.html

Edulampang. (2024, November). DE shows results of cracking down on online thieves for 1 year, opens AOC 1441 center, suspends 340,000 suspicious accounts - 19 billion baht in damages, cracks down seriously until statistics decrease. Retrieved from https://edulampang.prd.go.th/th/content/category/detail/id/57/iid/340064

ETDA. (2023). *Survey on online transactions behavior.* Electronic Transactions Development Agency (ETDA).

Google. (2025, April). NBTC partners with Google Cloud to enhance Thailand's cyber security through AI-driven cyber threat prevention collaboration. Retrieved from https://thailand.googleblog.com/2025/04/strategic-cybersecurity-collaboration-NCSA.html

Group-IB. (2024, February). iOS trojan stealing facial recognition data. Retrieved from https://www.group-ib.com/blog/goldfactory-ios-trojan/

Ministry of Digital Economy and Society. (2023, November). Minister of Digital Economy kicks off AOC 1441 center to solve online fraud problems as one stop service for the public. Retrieved from https://www.mdes.go.th/news/detail/7535

Nation Thailand. (2024, October). Over a quarter of Thais targeted by scams over past years. Retrieved from https://www.nationthailand.com/news/general/40042159.

Nation Thailand. (2024b, October). Thai Gov Tigthens Cybersecurity Laws. Retrieved from https://www.nationthailand.com/thailand/general/40032467

Nation Thailand. (2025a, February). Govt to step up online scam crackdown. Retrieved from https://www.nationthailand.com/news/general/40045195

Nation Thailand. (2025b). New policy measures to combat online fraud. Retrieved from https://www.nationthailand.com/news/policy/40045647.

Reuters. (2025, January). Thailand, China set up coordination centre to combat scam call networks. Retrieved from https://www.reuters.com/world/asia-pacific/thailand-china-set-up-coordination-centre-combat-scam-call-networks-2025-01-24.

Royal Thai Police. (2023). *Cyber Crime Investigation Bureau Media Release.* Royal Thai Police.

Royal Thail Government. (2025, February). AOC 1441 warns of “online thieves” who trick people into investing in stock trading - threatening to transfer money, claiming to be police, resulting in a loss of over 16 million baht. Retrieved from https://www.thaigov.go.th/news/contents/ministry_details/93747

Siam Legal. (2025, Febryary). What is Thailand’s Anti-Online Scam Operation Center? Retrieved from Siam Legal: https://library.siam-legal.com/what-is-thailands-anti-online-scam-operation-center/

The Irish News. (2025, February). Crackdown sees thousands of scam centre workers awaiting repatriation. Retrieved from https://www.irishnews.com/news/world/crackdown-sees-thousands-of-scam-centre-workers-awaiting-repatriation-6LAWDMNCLZLD5GUKX4RQGQWLC4/.

Tilleke & Gibbins. (2025). Thailand's new cybercrime measures enlist aid of banks and service providers. Retrieved from https://www.tilleke.com/insights/thailands-new-cybercrime-measures-enlist-aid-of-banks-and-service-providers/

UNODC. (2024). *Transnational organized crime and the convergence of cyber-enabled fraud, underground banking and technological innovation in Southeast Asia: A Shifting threat landscape.* UNODC. Retrieved from https://www.unodc.org/roseap/uploads/documents/Publications/2024/TOC_Convergence_Report_2024.pdf

Online Fraud and Scams in Vietnam

Nong Phuong Thao

11

Online Fraud and Scams in Vietnam

Nong Phuong Thao[24]

INTRODUCTION

In recent years, Vietnam has witnessed rapid digital transformation, driven by the expansion of e-commerce, digital banking, and widespread internet penetration. As the country embraces new technologies, the risks associated with cyber threats have also intensified, with online scams emerging as a significant concern. Among these, the rise of generative artificial intelligence (AI) has introduced a new dimension to cyber fraud, enabling scammers to create highly sophisticated and convincing schemes that exploit unsuspecting individuals and businesses. AI-powered tools allow cybercriminals to automate large-scale phishing attacks, develop deepfake videos for identity theft, and produce fraudulent websites that closely mimic legitimate ones. The evolving landscape of online scams threatens not only financial security but also public trust in Vietnam's digital ecosystem, posing challenges for the government, businesses, and individuals alike.

Vietnam's dynamic economic growth, coupled with its strategic position in the ASEAN region, makes it an attractive target for cybercriminals operating across borders. As one of the fastest-growing digital economies in Southeast Asia, the country has experienced a surge in cross-border online fraud incidents, where scammers leverage AI to exploit regulatory loopholes and jurisdictional challenges. The increasing reliance on digital platforms for financial transactions and communication further exacerbates the situation, making it imperative for Vietnam to adopt a proactive approach in addressing the threats posed by AI-driven scams. The regulatory and law enforcement mechanisms currently in place are often reactive rather than preventive, highlighting the need for a comprehensive strategy that integrates technological solutions with policy interventions.

Vietnam's government, in collaboration with international partners, has launched various initiatives to enhance cybersecurity awareness and strengthen its digital infrastructure. However, public awareness campaigns have not kept pace with the sophistication of AI-driven scams, leaving many citizens and small businesses vulnerable to cyber threats. The private sector, including banks and e-commerce platforms, has also taken steps to implement fraud detection systems powered by AI; however, the lack of collaboration and information-sharing between stakeholders remains a significant barrier to comprehensive fraud prevention efforts.

This report aims to provide an in-depth analysis of the current patterns and trends of AI-enabled online scams in Vietnam, offering insights into the evolving tactics used by cybercriminals and the vulnerabilities that exist within the country's digital infrastructure. It will assess the effectiveness of existing national and regional policies designed to combat online fraud, identifying gaps and areas for improvement. The scope of the policy assessment encompasses both domestic and cross-border aspects of online scams, recognizing that the problem is not confined within national borders. The research also seeks to evaluate the social and economic impact of such scams on businesses and individuals, with a particular focus on vulnerable populations, such as small and medium enterprises (SMEs) and individuals with limited digital literacy. Furthermore, this study will offer strategic recommendations aimed at strengthening Vietnam's regulatory frameworks, improving cross-border collaboration, and enhancing public awareness initiatives. By leveraging both domestic and international best practices, Vietnam can build a more secure and trustworthy digital environment, protecting its citizens and businesses from the ever-evolving threats of AI-enabled online fraud.

[24] Researcher, Central Institute for Economic Management (CEM)

FINDINGS AND ANALYSIS

Patterns and Trends

Types of scams

Online scams in Vietnam are an issue that has been and continues to receive significant attention from society. Malicious actors are taking advantage of the rapid explosion of information technology, including advancements in artificial intelligence (AI), to carry out numerous online fraud schemes, seizing high-value assets. The rise of generative AI and sophisticated AI tools has enabled scammers to automate and scale their operations, making scams more convincing and harder to detect.

There are three main types of online scams in Vietnam such as brand impersonation, account takeover, and various hybrid methods—with 24 specific scam tactics currently occurring in Vietnam's cyberspace (see Annex A). Scammers increasingly leverage AI-driven techniques, such as deepfake technology, AI-generated phishing emails, and voice cloning, to impersonate trusted individuals or organizations with an unprecedented level of realism. This has resulted in a growing number of victims falling prey to fraudulent schemes that exploit AI-generated content to bypass traditional security measures.

AI-powered scams, such as deepfake and deepvoice video call scams, represent a new frontier in online fraud. These technologies allow scammers to convincingly impersonate individuals, exploiting trust-based relationships to deceive victims. For instance, using deepfake videos and voice cloning, criminals can impersonate a family member or colleague in real-time video calls, making their fraudulent requests seem authentic. Similarly, phishing links and deceptive advertisements on online platforms are increasingly powered by AI algorithms that analyze user behavior to target victims more effectively. These scams demonstrate the transformative power of AI in amplifying the scale and sophistication of cybercrime in Vietnam. Another noteworthy example is the "FlashAI" scam, which leverages AI-generated fake calls or messages to manipulate victims into believing they are losing money. The psychological pressure created by such real-time and personalized interactions significantly increases the success rate of these scams. These methods highlight the evolving nature of fraud, where AI enables precise targeting and realistic simulations that traditional methods cannot achieve.

Despite the rise of AI-driven fraud, traditional scams remain a significant concern in Vietnam. These include tactics like impersonating financial institutions, forging money transfer receipts, and creating fake job opportunities. Such scams often exploit low digital literacy, trust in official institutions, and lack of robust verification mechanisms. For example, scams involving fake SIM card lock warnings or counterfeit e-commerce platforms capitalize on users' unfamiliarity with digital safety practices. While these methods lack the technological sophistication of AI-driven scams, they are still effective due to their simplicity and adaptability to various contexts.

Targeted victims

Scammers today employ a wide array of increasingly sophisticated techniques, leveraging both traditional and AI-driven methods to target two key categories: individuals and organizations/ businesses. This segmentation allows scammers to tailor their methods based on the vulnerabilities, behaviors, and digital footprints of their targets.

a) Individual targets

With AI-powered data analysis and social engineering tactics, scammers can personalize their attacks based on an individual's online behavior, preferences, and vulnerabilities. For each age

group, scammers use different AI-driven tactics to lure their victims, such as chatbot-based fraud schemes or AI-generated fake news to manipulate trust (see Annex B). The common goal remains to gain trust, steal personal information, and ultimately misappropriate assets.

AI-driven scams, such as those involving Deepfake video calls, demonstrate how malicious actors leverage advanced technologies to enhance the believability of their schemes. In Vietnam, where digital transformation is progressing rapidly, yet digital literacy remains uneven across demographic groups, the elderly are particularly vulnerable, as they often lack awareness of how AI technologies can make fraudulent calls appear convincingly authentic.

Meanwhile traditional scams, such as fake brand promotions, phishing, and fraudulent online transactions, still dominate the landscape. These scams often exploit trust in familiar institutions, such as government agencies or e-commerce platforms. Their prevalence highlights gaps in public awareness and digital security practices, particularly among the elderly and youth. While these scams do not rely on AI, their continued success underscores the importance of education campaigns to raise awareness and build resilience against basic fraud tactics.

The table also illustrates how scammers tailor their tactics to specific demographic vulnerabilities. The elderly, for example, are targeted with scams impersonating government or financial institutions, leveraging their perceived trust in authority. Meanwhile, youth and students are more likely to face recruitment scams and fraudulent apps that exploit their familiarity with digital tools but limited experience with cyber threats. The use of AI-driven Deepfake calls in both groups shows the versatility of AI in enabling scams to penetrate diverse segments of the population.

Scammers exploit the financial insecurity of low-income workers and the limited cybersecurity awareness in rural communities by leveraging AI tools like chatbots, personalized phishing emails, and AI-generated content to create false legitimacy. For low-income workers, schemes such as Ponzi scams and fraudulent investment platforms use fake reviews and user-generated content to promise quick financial gains. Similarly, rural communities, often reliant on mobile devices and lacking access to cybersecurity education, are targeted with AI-generated fake job offers and fraudulent online lending platforms tailored to their financial aspirations.

b) Organizational/Business Targets

Small and medium-sized enterprises (SMEs), are also a major focus for scammers. SMEs are particularly vulnerable due to their limited cybersecurity resources and reliance on digital communications. AI-powered phishing attacks and Business Email Compromise (BEC) scams are among the most common tactics used. These involve highly personalized emails that mimic business partners or clients, often containing convincing details like invoice numbers or account specifics. Such scams deceive SME owners or employees into transferring funds or sharing sensitive information. Another prevalent scam targeting SMEs involves fake invoices and fraudulent transactions. Scammers use AI to generate realistic fake invoices, exploiting SMEs' reliance on electronic payment systems. This can lead to significant financial losses for businesses operating on tight margins.

In addition, financial institutions, telco industry, and e-commerce platforms, faces mounting challenges in combating AI-powered fraud. Banks such as TPBank and Techcombank report a surge in cases where scammers use AI-driven deepfake technology and voice cloning to bypass biometric authentication systems. Fraudsters have manipulated eKYC (electronic Know Your Customer) verification processes, allowing them to create fraudulent bank accounts under stolen

identities. A notable case involved a Techcombank customer who lost 14.6 billion VND after scammers posed as law enforcement officers and used AI-generated voice calls to coerce the victim into transferring funds. The telecommunications sector has also noted an increase in AI-enhanced SMS phishing ("smishing") attacks, where scammers deploy fake base transceiver stations (BTS) to send fraudulent messages impersonating government agencies or financial institutions.

Emerging trends – AI-driven scams

Several key trends highlight the increasing scale and impact of AI-driven scams, including widespread data leaks, AI-powered identity theft, deepfake impersonations, and automated scam operations. These trends pose severe risks to individuals, businesses, and national security, necessitating urgent regulatory and technological interventions.

a) AI-Driven Data Leaks and Personal Information Exploitation

One of the most alarming trends in online fraud in Vietnam is the escalation of personal data leaks, which serve as the foundation for increasingly sophisticated AI-driven scams. According to the National Cyber Security Center (NCSC) under Ministry of Public Security (2024), cybercriminals primarily acquire personal information from underground marketplaces, where data is bought and sold via chatbots and paid for with cryptocurrency, making transactions difficult to trace. The number of individual accounts compromised in 2024 reached 121,482,341, marking a 15.8% increase compared to 2023 (104,917,940 accounts). The table below shows some of the sectors with most affected accounts:

Table 11.1 Number of Affected Accounts Based on Sectors

Sectors	Number of affected accounts
Public services	5,271,054
Social media	5,905,760
Banking and digital payment	1,161,713
Healthcare and education	335,314

The root causes of these data leaks are multifaceted. NCSC also indicates that 73.99% of users attribute leaks to providing information during online shopping, while 62.13% believe due to sharing data on social media. Additionally, 67% of users reported data leaks from essential services such as restaurants, hotels, and supermarkets, where information security measures are often weak. These breaches enable scammers to create highly personalized scams using AI-powered analytics, increasing the likelihood of victims falling for fraud attempts.

b) Escalation of AI-Enhanced Phishing and Social Engineering Scams

With access to vast amounts of stolen personal data, cybercriminals are using AI to craft hyper-personalized phishing campaigns. Traditional phishing emails were often generic and riddled with grammatical errors, making them easier to detect. However, modern scams leverage AI-powered natural language processing (NLP) algorithms to generate flawless, contextually relevant phishing messages. These scams frequently mimic official communications from banks, government

agencies, and e-commerce platforms, tricking victims into revealing sensitive information or making financial transfers. According to NCSC, an increasing 66.24% of users in Vietnam report that their personal information has been used illegally, demonstrating the effectiveness of AI-driven scams. Furthermore, 1 in every 220 smartphone users in Vietnam falls victim to fraud, with financial investment scams, impersonation fraud, and fake lottery winnings among the most prevalent. Alarmingly, despite the large number of victims (70% of Vietnamese encounter scam attempts at least once a month), only 45.69% report scams to authorities, making law enforcement efforts even more challenging. In addition, the fact that only 1% of victims have successfully recovered their stolen money shows the urgent need for stronger frameworks to address scams.

According to the Vietnam Scam Report 2023, AI-powered phishing attacks are becoming harder to detect due to their linguistic accuracy and personalized nature. Cybercriminals leverage AI-driven chatbots and automated email generators to create hyper-realistic messages that mimic banks, law enforcement agencies, or e-commerce platforms. These messages often include deepfake voice recordings or AI-generated images to further deceive victims. Although phone calls and SMS remain the primary scam outreach channels, social media platform and email are also growing targets.

c) Automated Scam Call Networks and Chatbot-Assisted Fraud

The use of AI-generated deepfakes and synthetic voices is reshaping impersonation scams in Vietnam. Fraudsters are now able to clone voices and faces of family members, government officials, or bank representatives, making phone scams almost indistinguishable from legitimate calls. For instance, 62.08% of Vietnamese users reported receiving scam calls impersonating police, tax agencies, and banks, urging them to install software or transfer money under false legal threats. While 60.01% received fake prize notifications, often accompanied by AI-generated promotional videos.

These AI-powered deepfakes are being used to gain victims' trust quickly, persuading them to divulge sensitive information or approve unauthorized transactions. The Trust system, an anti-fraud initiative, has recorded 134,000 reports of scam phone numbers in just six months, forcing authorities to continuously update a list of fraudulent numbers, which reached 296,000 in 2024.

AI-driven chatbots are also being deployed on social media and e-commerce platforms to engage victims in real-time. These bots can simulate human conversations, answer queries, and build trust, leading victims to share financial details or make fraudulent transactions. Given Vietnam’s high social media penetration, these scams have been particularly effective in targeting younger demographics who frequently engage in online commerce.

d) Rise of AI-Enhanced Malware and Ransomware Attacks

The increasing reliance on digital platforms for work, finance, and social interactions has exposed Vietnamese users to sophisticated AI-enhanced malware attacks. According to a survey by the National Cyber Security Association, 23.4% of users reported being attacked by malware at least once a year and 9.65% suffered financial and data losses. These malware variants can steal login credentials, hijack financial accounts, or lock users out of their own devices. The risk is amplified by the widespread use of personal devices for both work and leisure, creating new vulnerabilities for corporate data breaches.

Socio-Economic Implications of Scams

The persistent threat of online scams has broader implications for Vietnam's economic development and technological advancement. As the country aspires to become a regional hub for digital innovation and e-commerce, continued incidents of fraud pose a significant risk to Vietnam's reputation in the global market. Businesses and investors may hesitate to expand their operations in Vietnam if they perceive the digital environment as unsafe or unregulated. Without decisive action to curb online scams, Vietnam risks undermining public confidence in its digital economy, slowing down its progress toward Industry 4.0 and broader economic integration.

Economic impacts

Online scams in Vietnam have resulted in substantial economic losses, affecting individuals, businesses, and financial institutions. The financial impact of online fraud in Vietnam has more than doubled in just one year. In 2023, online fraud caused losses of 8,000 billion VND, but by 2024, this figure surged to 18,900 billion VND, a 2.36 times increase. The economic burden is not just limited to direct monetary losses; businesses also suffer from brand damage, loss of consumer trust, and the costs associated with fraud prevention. The banking sector, e-commerce platforms, and telecommunications companies bear the brunt of these attacks, often having to refund stolen funds or implement expensive security upgrades to protect their customers. In addition, foreign investors are increasingly concerned about the potential risks associated with Vietnam's growing cybercrime landscape, potentially affecting the country's attractiveness as a business destination.

Social consequences

Beyond the economic ramifications, AI-driven online scams have significant social implications for Vietnamese society. A rising number of scam victims experience emotional and psychological distress, including anxiety, depression, and a loss of confidence in digital interactions. Scammers often exploit vulnerable populations such as elderly citizens and individuals with low digital literacy, leaving them feeling isolated and distrustful of technology. The proliferation of online scams also erodes public trust in digital platforms, slowing down Vietnam's digital transformation goals by discouraging users from engaging in e-commerce, online banking, and other digital services. Families of scam victims may experience strained relationships due to financial losses and social stigma associated with falling victim to fraud. In addition, the widespread fear of online fraud has led to a growing demand for government intervention and stricter regulations, increasing pressure on policymakers to implement robust cybersecurity measures while balancing economic growth and digital innovation.

Vulnerable groups

Certain demographic groups in Vietnam are disproportionately affected by online scams, exacerbating existing social and economic inequalities. Elderly individuals, who may have limited experience with digital platforms, are particularly susceptible to AI-driven fraud schemes such as phishing emails, fraudulent investment opportunities, and impersonation scams. Similarly, rural populations with lower access to digital literacy programs are at high risk of falling victim to scams, as they often lack awareness of cyber threats and preventive measures. Additionally, low-income individuals who seek online financial opportunities or loans are frequently targeted by scammers offering fake job opportunities or financial assistance, further entrenching them in financial distress. The impact of these scams on vulnerable groups not only results in personal financial hardship but also deepens the digital divide, limiting the potential benefits of Vietnam's push toward a digital economy.

STRATEGY IN ADDRESSING ONLINE FRAUD AND SCAMS

The Role of Stakeholders in Addressing Online Fraud and Scams

Several organizations, such as government, private sectors and CSOs play an important role in addressing the challenges derived from the scams. The table below shows some of the challenges and initiatives that has been taken in Vietnam.

Table 11.2 Initiatives From Key Institutions to Address Scams

Type of Institutions	Institutions/ Organizations	Initiatives and Challenges
Government	The Ministry of Public Security (MPS)	Leads law enforcement efforts by investigating fraud cases, cracking down on cybercrime, and enhancing international cooperation.
	The Ministry of Planning and Investment (MPI)	Strengthens business registration regulations to prevent fraudulent companies from operating.
	Ministry of Finance (MOF)	Oversees financial markets, ensuring transparency and preventing fraudulent investment schemes.
	The State Bank of Vietnam (SBV)	Regulates digital payments, preventing financial fraud, and coordinating with law enforcement to block suspicious transactions.
	Ministry of Industry and Trade (MOIT)	Monitors e-commerce activities, cracking down on fake online stores and illegal multi-level marketing (MLM) schemes.
	The Ministry of Information and Communications (MIC)	Enhances cybersecurity regulations, collaborates with tech companies to remove fraudulent content, and promotes digital literacy among consumers.
	the Supreme People's Court and Supreme People's Procuracy	Ensure strict judicial enforcement, prosecuting offenders and securing asset recovery for victims.
Private Sector	Banks and financial institutions	Implemented biometric security measures, AI-powered fraud detection, and multi-factor authentication (MFA) to strengthen customer verification processes.

		Integrated bio-authentication, NFC scanning for chip-embedded IDs, and transaction monitoring via Big Data analytics.
	Telco	Enhanced spam filtering systems, implemented real-time monitoring of network anomalies, and worked closely with law enforcement to detect fraudulent SIM card activations.
Civil society organizations (CSOs)		Organized digital literacy campaigns, published scam alerts, and provided legal consultation services for scam victims. Researching AI-driven scams, identifying vulnerabilities in cybersecurity frameworks, and developing fraud detection technologies.

POLICY OVERVIEW

National Policies

Vietnam has made significant strides in developing legal frameworks to address cybersecurity and online fraud; however, the current regulations fall short in tackling the emerging challenges posed by generative AI in online scams, including:

Table 11.3 Vietnam National Policies to Address Scams

Type of policies	Name of Regulations/Initiatives	Description
Regulations	Law on Cyber Security (2018) and Decree 53/2022/ND-CP	Provides a legal framework for managing and protecting cyberspace, preventing cybercrime, including transnational online fraud
	Criminal Law (2015, amended in 2017)	Article 174 stipulates the crime of fraud and appropriation of property
	Law on Electronic Transactions (2005)	Protects consumer rights in the digital environment, including international transactions
	Law on Information Technology (2006)	Engagement in information technology application and development activities in Vietnam

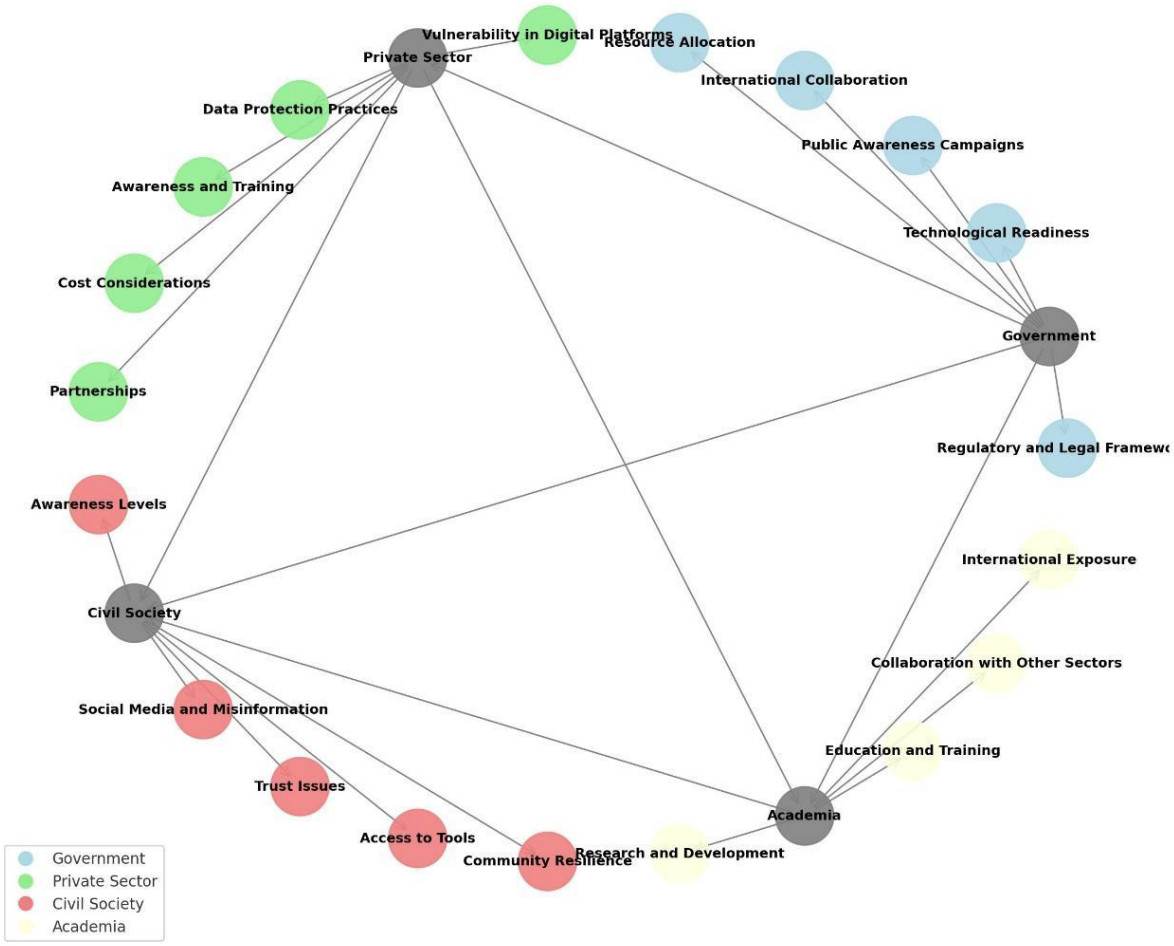
	Law on Protection of Consumer Rights (2023)	principles and policies for protecting consumers' rights
	The Directive No. 21/CT-TTg (2020)	government's commitment to strengthening the prevention and handling of fraudulent activities related to asset appropriation
	• Decree No. 13/2023/ND-CP on personal data protection. • Decree No. 116/2013/ND-CP on Anti-Money Laundering. • Decision No. 2345/QD-NHNN on safety and security measures to online payment and card payment	Regulates to meet data protection requirements, the management, provision, and use of Internet services, including measures to prevent online fraud
	Decision 1811/QD-BTTTT on the plan strengthen management and minimize abuse of domain names to violate the law	Implement comprehensive measures to prevent and address domain name abuse, particularly for international and cross-border domains, while raising public and business awareness about safe and reliable websites through the promotion of the national domain ".vn."
Initiatives	The National Cyber Security Center (NCSC)	Responsible for monitoring and ensuring cybersecurity across Vietnam's cyberspace, directing telecommunications and Internet service providers, collecting and analyzing cyber threats, and issuing early warnings, and facilitates information sharing between domestic and international organizations to support regulatory enforcement.
	Hotline of the Department of Cyber Security and High-Tech Crime Prevention at 069.219.4053	Hotline for citizens and businesses can report cybersecurity incidents or seek assistance, managed by the Criminal Police Department
	Vietnam Information Security Warning Page at https://canhbao.ncsc.gov.vn	Access real-time alerts and resources which provides updates, guidance, and preventive measures to enhance cybersecurity awareness and response

Vietnam has established a comprehensive legal framework to address cybersecurity and online fraud through various laws and regulations. These legal instruments provide a solid foundation for managing and protecting cyberspace, ensuring social order, and prosecuting cybercriminals, including those involved in transnational fraud. However, the current legal framework primarily focuses on traditional cyber threats and lacks specific provisions to tackle the emerging risks posed by generative AI technologies. Advanced threats such as AI-generated phishing, deepfake fraud, and automated identity

theft remain unaddressed, leaving gaps in the regulatory landscape. As AI-driven scams become increasingly sophisticated, Vietnam's existing policies must evolve to include targeted measures that can effectively counter these challenges. There is an urgent need to incorporate AI-specific risk assessments and develop advanced detection systems capable of identifying AI-generated threats in real time. In addition, the complexity of this issue demands a multi-stakeholder approach involving the government, private sector, civil society, and academia.

Moreover, Vietnam's cybersecurity policies do not adequately address the cross-border nature of AI-driven fraud. Generative AI technologies transcend national boundaries, making it essential for Vietnam to strengthen its international cooperation efforts. Aligning domestic policies with global best practices and participating in international cybersecurity frameworks will be crucial to tackling AI-related cyber threats. Collaborative efforts with international cybersecurity organizations, regulatory harmonization, and the exchange of intelligence will enhance Vietnam's ability to combat transnational AI-driven fraud and bolster its cybersecurity capabilities on a global scale.

Figure 11.1 Factors Driving the Interaction of AI-Driven Online Scams in Vietnam



Source: author's compilation

Cross-Border Measures

The rapid advancement of generative AI has significantly escalated the sophistication and scale of cross-border online scams targeting Vietnamese citizens and businesses. According to the survey, fraudsters are exploiting cryptocurrency transactions and anonymous payment channels, contributing to an estimated $37 billion stolen in Southeast Asia in 2023, with Vietnam accounting for a significant portion. One of the most alarming cases in recent years involved a Cambodia-based fraud ring that used deepfake technology to mimic Vietnamese police officers and tax officials, coercing victims into transferring funds under false pretenses. This operation defrauded over 13,000 individuals, causing financial losses estimated at 1,000 billion VND (Vietnamnet, 2025).

Vietnam's fight against AI-driven cross-border scams is particularly challenging due to the international nature of these criminal networks. Law enforcement efforts are further hindered by the fact that many of these scams operate outside Vietnam's jurisdiction, making it extremely difficult to arrest key perpetrators or recover stolen funds. Investigations have revealed that most large-scale scams affecting Vietnam are operated in Cambodia, led by foreign nationals, and using offshore bank accounts to obscure their financial trails.

Despite Vietnam's engagement in regional cybersecurity initiatives, including cooperation with ASEAN and INTERPOL, gaps remain in enforcement capabilities. Existing legal frameworks in Southeast Asia are not fully harmonized, creating loopholes that cybercriminals exploit. The lack of standardized AI governance policies across ASEAN countries further complicates efforts to track and dismantle AI-driven fraud networks. Even when suspects are arrested, the use of offshore accounts and decentralized cryptocurrency transactions makes financial recovery nearly impossible.

Recognizing these threats, Vietnamese authorities have ramped up efforts to dismantle AI-enhanced fraud networks. A recent joint operation between multiple provincial police departments and national agencies successfully disrupted a Cambodia-based deepfake scam operation, leading to the arrest of 60 individuals. However, these enforcement actions remain reactive rather than preventive, highlighting the need for stronger intelligence-sharing mechanisms, AI-driven fraud detection systems, and cross-border policy alignment.

POLICY GAPS IN EFFECTIVELY ADDRESSING SCAMS

Despite Vietnam's ongoing efforts to strengthen cybersecurity and combat online scams, several critical gaps remain in addressing the growing threat posed by AI-driven fraud. These gaps span across regulatory frameworks, enforcement capabilities, inter-agency coordination, public awareness, and cross-border cooperation. The rapid evolution of generative AI has outpaced the country's current policies and detection mechanisms, making it increasingly difficult for authorities and private entities to effectively combat these sophisticated scams.

Narrow Coverage of the Existing Cyber Law

The emergence of deepfakes and AI-driven scams poses a significant challenge for governments in tackling these evolving threats. Although existing laws such as the Cybersecurity Law (2018) and the Personal Data Protection Decree (2023) provide a foundation for regulating online fraud, it does not yet encompass the complexities of AI-powered cyber threats. These laws are primarily designed to address conventional cybercrimes such as hacking and identity theft. Thus, revisit and revise the current Cybersecurity Law (2018), expanding its scope to redefine fraud and scams, and explicitly include deepfakes within its provisions is needed to address fraud and AI-driven scams.

Enforcement Limitations and Technical Capacity Constraints

Even when online scams are identified, Vietnamese law enforcement agencies face significant challenges in investigating and dismantling these operations. A major obstacle is the limited technical expertise and insufficient resources available to combat AI-driven fraud effectively. Cybercriminals frequently operate from foreign jurisdictions, using encrypted communications, offshore accounts, and decentralized payment systems, making it nearly impossible for domestic agencies to track, identify, and arrest perpetrators.

The Ministry of Public Security (MPS), which oversees cybercrime investigations, often lacks the advanced AI-powered forensic tools needed to analyze deepfake content, detect AI-generated phishing attacks, or trace funds linked to fraudulent activities. Additionally, there is a shortage of trained personnel specializing in AI-related cybersecurity threats, which weakens Vietnam's ability to stay ahead of cybercriminal tactics. Compared to leading cybersecurity agencies in countries like Singapore or South Korea, Vietnam's law enforcement lacks real-time threat detection capabilities and AI-driven fraud monitoring systems, leaving them in a reactive position rather than proactively preventing scams.

Fragmented Coordination Among Regulatory Bodies

Vietnam's cybersecurity efforts are hampered by fragmented coordination between government agencies, financial institutions, and digital service providers. The Ministry of Information and Communications (MIC), the Ministry of Public Security (MPS), and the State Bank of Vietnam (SBV) all play roles in cybersecurity governance and fraud prevention, yet they operate under separate mandates without a unified approach. This lack of inter-agency collaboration results in delays in responding to scams, inefficient information sharing, and inconsistent enforcement actions. For example, financial institutions and e-commerce platforms frequently detect fraudulent activities but struggle to coordinate with law enforcement agencies due to unclear reporting protocols. Additionally, there is no centralized AI fraud monitoring system where financial institutions, telecom companies, and government bodies can share real-time scam alerts and intelligence.

To effectively combat AI-driven scams, Vietnam needs to establish an integrated fraud response framework that facilitates information sharing between government agencies, financial institutions, e-commerce platforms, and telecom providers. This would allow for faster detection, improved collaboration, and more coordinated enforcement efforts.

Low Public Awareness and Digital Literacy

Another major gap in Vietnam's fight against AI-driven online scams is the lack of widespread public awareness and low levels of digital literacy, especially among vulnerable populations. Many Vietnamese citizens, particularly the elderly, rural communities, and small business owners, are unaware of how AI is being used in online fraud. Scammers frequently exploit low cybersecurity awareness, using AI-generated voice calls, fake government notifications, and deepfake videos to impersonate trusted authorities or financial institutions.

Although some public awareness campaigns have been launched by government agencies and financial institutions, they have not kept pace with the sophistication of AI-enhanced scams. Many victims remain unaware of how to verify digital identities, detect AI-generated fraud, or report scams to the appropriate authorities. A recent survey found that only 45.69% of scam victims reported their cases to law

enforcement, demonstrating a lack of trust in the system and insufficient knowledge of reporting mechanisms.

To address this, Vietnam needs to implement nationwide digital literacy programs, particularly in rural areas and among high-risk groups. Public awareness campaigns should focus on educating users about AI-generated phishing, deepfake fraud, and common scam tactics, while financial institutions and telecom providers should integrate fraud prevention alerts into their customer communication strategies.

Challenges in Cross-Border Collaboration

Vietnam's struggle with cross-border AI-driven scams is further exacerbated by inconsistent international cooperation and legal limitations in tracking cybercriminal networks operating beyond its borders. Many large-scale scams affecting Vietnam originate from Cambodia, China, and other Southeast Asian countries, where fraud syndicates operate call centers and AI-powered scam factories targeting Vietnamese citizens. These operations use foreign bank accounts, cryptocurrency transactions, and fake identities, making it nearly impossible for Vietnamese law enforcement to recover stolen funds.

While Vietnam has participated in ASEAN cybersecurity initiatives and collaborated with INTERPOL, these partnerships have not translated into effective cross-border enforcement mechanisms. Jurisdictional barriers, differences in legal frameworks, and the lack of formal extradition agreements make it difficult for Vietnamese authorities to prosecute foreign-based scammers. Additionally, regional law enforcement agencies lack AI-specific expertise, further limiting their ability to investigate AI-driven cybercrime effectively.

To improve cross-border enforcement, Vietnam should push for greater legal harmonization across ASEAN, advocate for bilateral agreements with key regional partners, and work with international cybersecurity organizations to establish a dedicated AI fraud intelligence-sharing network. Strengthening cross-border financial tracking mechanisms and enhancing collaboration with global tech companies would also help in identifying scam operators and shutting down fraudulent operations.

POLICY RECOMMENDATIONS

Areas of Action by Government

The Vietnamese government plays a central role in developing and enforcing policies to mitigate the risks posed by AI-driven scams. As AI fraud techniques continue to evolve, Vietnam's legal framework and enforcement capabilities must adapt swiftly to protect individuals and businesses from cybercrime. The government must focus on strengthening regulations, enhancing enforcement mechanisms, fostering regional cooperation, and increasing public awareness.

Strengthening Legal and Regulatory Frameworks

Vietnam's current cybersecurity and fraud prevention laws are outdated when it comes to addressing AI-driven scams. The Cybersecurity Law (2018), which governs Vietnam's online security framework, does not explicitly cover AI-generated deepfake fraud, voice phishing scams, or synthetic identity fraud. Without precise legal definitions, law enforcement agencies face challenges in tracking and prosecuting cybercriminals who exploit AI tools.

To address this issue, the government must reviews the existing laws, and incorporating AI-powered scams into the legal definitions of fraud and scams. The Penal Code (2015, amended 2017) should be

expanded to include harsh penalties for individuals and organizations using AI technologies for fraudulent activities. Additionally, mandatory reporting mechanisms should be enforced for financial institutions, telecom providers, and digital platforms, requiring them to report suspected AI-driven scams to law enforcement agencies in real time.

A National AI Governance Framework should be introduced, outlining ethical AI usage, responsible AI development, and enforcement measures against misuse. This framework should align with international AI security standards, such as the OECD AI Principles and the EU AI Act, ensuring that Vietnam's AI governance is in line with global best practices.

Institutional Capacity Building and Enforcement Enhancement

Law enforcement agencies, including the Ministry of Public Security (MPS) and the Cybersecurity and High-Tech Crime Prevention Department (A05), are in the need for advancing capacity in AI-driven scams including strengthen specialized training and advanced forensic tools. The government is recommended to establish a specialized AI Fraud Detection Unit, equipped with cutting-edge AI forensic tools that can analyze fraudulent transactions, detect deepfake scams, and track AI-generated phishing campaigns.

To build expertise, comprehensive training programs must be implemented for law enforcement officers, prosecutors, and judges, focusing on AI-powered cybercrime investigation, digital forensics, and international cybersecurity standards. These efforts will enable authorities to track, investigate, and prosecute AI-enhanced fraud more effectively. Additionally, the deployment of AI-powered surveillance systems is crucial for real-time fraud detection. Government agencies should partner with AI firms and cybersecurity organizations to integrate machine learning algorithms that can identify fraudulent activities across social media, financial transactions, and telecom networks.

Public Awareness and Consumer Protection

AI-driven scams thrive on public ignorance and low digital literacy. Many victims fall prey to deepfake impersonation scams, AI-generated phishing emails, and fake investment schemes due to a lack of awareness.

To combat this, the government should launch a nationwide Digital Scam Awareness Program, targeting individuals, small businesses, and vulnerable groups such as elderly citizens and rural communities. This program should include mass media campaigns, online tutorials, and school-based cybersecurity education programs. Government agencies should also collaborate with banks, telecom providers, and social media platforms to create fraud alerts and scam detection tools that warn users about AI-generated fraud attempts in real time.

Strengthening Cross-Border Collaboration

Since many AI-driven scams in Vietnam originate from international cybercrime networks, Vietnam must prioritize cross-border cybersecurity collaboration. The lack of regional cooperation has allowed cybercriminals to exploit legal loopholes between ASEAN nations, making it difficult to track and prosecute foreign-based scam operators.

Vietnam should advocate for an ASEAN-wide AI Scam Prevention Task Force, which would facilitate real-time intelligence sharing, coordinated law enforcement operations, and policy harmonization among ASEAN countries. Vietnam should also expand its cooperation with INTERPOL, APEC, and global cybersecurity alliances, enabling faster extradition agreements and joint cybercrime investigations.

A critical challenge in tackling cross-border AI fraud is the use of offshore bank accounts and cryptocurrency transactions to launder stolen funds. The Vietnamese government must work with regional financial regulators to establish cross-border financial fraud monitoring mechanisms that detect suspicious transactions linked to AI-driven scams.

Areas of Action by Private Sector

The private sector, including financial institutions, e-commerce platforms, telecommunications companies, and technology firms, is at the frontline of detecting and preventing AI-driven online scams. As fraudsters increasingly exploit generative AI for phishing, deepfake impersonation, and automated scams, businesses must strengthen fraud detection systems, data security, consumer awareness, and industry collaboration to mitigate risks.

Adoption of Advanced Fraud Detection Technologies

One of the most critical steps in combating AI-driven online scams is investing in advanced AI security solutions. Financial institutions and e-commerce platforms should integrate AI-driven fraud detection tools capable of identifying deepfake-assisted scams, synthetic identity fraud, and phishing attacks. Banks should enhance biometric authentication, voice recognition, and behavioral analytics to detect fraudulent transactions in real time. Additionally, telecommunications firms must deploy AI-based call and SMS filtering systems to block voice-cloned scam calls and smishing attacks before they reach users.

Another essential measure is implementing real-time monitoring and risk scoring systems. Banks like TPBank and Techcombank have already introduced risk-based authentication and transaction monitoring tools that assess customer behaviors to flag unusual activities. This approach should be expanded across industries to include e-commerce and fintech sectors. Companies should also adopt automated scam detection algorithms capable of identifying fraudulent activities across digital platforms, banking systems, and payment gateways to prevent scams before they occur.

Enhancing Cross-Industry and Public-Private Collaboration

A major challenge in combating AI-driven scams is the lack of real-time fraud intelligence sharing among financial institutions, telecom providers, and digital platforms. To address this, the private sector should work toward developing a National AI-Driven Fraud Intelligence Network, allowing businesses to share data on fraudulent accounts, suspicious transactions, and scam-related content. A private sector consortium focused on fraud prevention could facilitate this exchange and improve collective defense mechanisms.

In addition, stronger public-private partnerships on cybersecurity are essential. The Ministry of Public Security (MPS), State Bank of Vietnam (SBV), and Ministry of Information and Communications (MIC) should collaborate with private companies to develop a centralized fraud detection platform. AI security firms, financial institutions, and digital platforms should participate in government-led cybersecurity exercises to test fraud prevention measures and refine AI-based scam detection models.

Consumer Awareness and Support

Consumer education is key to preventing AI-driven scams. Banks, e-commerce platforms, and telecom providers should launch fraud awareness campaigns via SMS alerts, emails, and mobile app notifications, warning users about AI-generated scams. Businesses should integrate AI-driven chatbots to assist customers in identifying fraudulent messages and transactions. Hosting online workshops and training sessions can further help users recognize deepfake impersonations and phishing attempts, especially targeting elderly individuals and rural communities, who are more vulnerable to scams. To improve fraud

reporting, companies must establish dedicated hotlines, online portals, and in-app reporting tools for users to quickly report suspicious activities, ensuring faster response and mitigation.

Strengthening User Protection Measures

To enhance security for users, financial institutions and e-commerce platforms must improve customer authentication and data security mechanisms. Banks should strengthen e-KYC (Electronic Know Your Customer) processes by integrating chip-embedded ID verification, biometric authentication, and AI-driven identity validation. Multi-factor authentication (MFA) should become mandatory for high-value transactions, particularly in cross-border payments, to reduce fraud risks. Additionally, e-commerce platforms should enforce stricter seller verification processes to prevent fraudulent listings and scams.

Alongside improved security measures, consumer education on AI-driven scams is crucial. Banks, e-commerce platforms, and telecom companies should organize nationwide digital literacy campaigns to help users recognize deepfake scams, phishing websites, and fraudulent job postings. Raising awareness about AI-generated fraudulent investment schemes and impersonation scams will empower consumers to identify potential risks and avoid falling victim to scams.

Addressing Cross-Border AI-Driven Fraud Challenges

As AI-driven fraud often operates across national borders, stronger regional cybersecurity agreements are necessary to address these threats. Vietnamese financial institutions and telecom providers should support the development of regional AI-driven fraud monitoring systems in partnership with ASEAN countries. International cooperation between law enforcement, financial regulators, and technology firms is crucial to tracking and blocking cross-border scam transactions.

At the same time, AI regulation and policy compliance must be strengthened to ensure that fraud prevention measures align with data protection laws. Companies must actively engage in shaping AI governance regulations to ensure that fraud detection frameworks are both effective and compliant with privacy regulations. Banks, in particular, should follow Vietnam's Personal Data Protection Decree (Decree 13) while ensuring that AI-driven fraud detection tools are used transparently and responsibly.

Improving Corporate Incident Response and Crisis Management

The private sector should also focus on establishing AI-powered scam incident response teams to detect and respond to AI-generated fraud cases swiftly. Financial institutions and digital platforms should have dedicated cybersecurity response units that can implement automated fraud alerts, emergency account freezes, and scam dispute resolution mechanisms. These teams would play a key role in minimizing financial losses and preventing further victimization.

Additionally, regular security audits and AI model testing must become a standard practice for private sector organizations. Banks, fintech firms, and e-commerce companies should conduct regular assessments of their AI-driven fraud detection models to ensure they remain effective. AI-based scam detection systems should undergo continuous testing against evolving fraud tactics to prevent fraudsters from adapting and circumventing security measures.

Areas of Action by Civil Society and Academia

As AI-driven scams continue to evolve in complexity, civil society organizations (CSOs) and academic institutions play a crucial role in shaping policies, raising awareness, and conducting research to mitigate the risks associated with these cyber threats. Their involvement ensures a balanced approach that complements government regulations and private sector initiatives by focusing on public education,

victim support, policy advocacy, and cybersecurity research. By strengthening collaboration between policymakers, businesses, and communities, CSOs and academia can help Vietnam build resilience against AI-enabled fraud. CSOs should play a more role in (i) promoting research in understanding the AI for fraudulent activities, (ii) collaborating with other stakeholders in developing digital literacy and cybersecurity programs for all groups, (iii) creating support services platforms for scam's victim, (iv) forming multistakeholder dialogue to bridge the research and policies, and (v) monitoring and evaluating anti-scams efforts.

ANNEXES

Annex A. Summary of 24 Specific Scam Tactics in Vietnam

No	Scam Type	Description	AI Technology/ Traditional
1.	Deepfake, Deepvoice video call scam	Scams using AI-generated voices and videos to impersonate individuals.	AI Technology
2.	Cheap travel package scam	Fraudulent offers of low-cost travel packages.	Traditional
3.	SIM card lock scam	Claiming a SIM card will be locked due to incomplete subscriber verification.	Traditional
4.	Fake successful money transfer receipt	Scammers forge receipts to show false transactions.	Traditional
5.	Impersonating teachers/medical staff	Scammers claim a relative is in an emergency to extort money.	Traditional
6.	Child model recruitment scam	Fraudsters lure victims with fake model job offers.	Traditional
7.	Financial institution impersonation	Scammers pose as banks or financial companies.	Traditional
8.	Gambling, betting, and loan app scams	Fraudulent apps and links promoting illegal gambling and loans.	Traditional
9.	Fake websites of institutions and businesses	Impersonation of official websites (e.g., social insurance, banks).	Traditional
10.	SMS brand name scam	Distribution of fraudulent SMS messages.	Traditional
11.	Stock, cryptocurrency, and Ponzi scheme scams	Fake investment opportunities promising high returns.	Traditional
12.	Online collaborator recruitment scam	False job offers for online work.	Traditional
13.	Social media account hacking	Stealing accounts to send scam messages.	Traditional
14.	Law enforcement impersonation scam	Fraudsters posing as police, prosecutors, or courts.	Traditional
15.	Selling counterfeit goods on e-commerce platforms	Fake products sold online.	Traditional
16.	Identity theft for credit loans	Using stolen ID cards to take out loans.	Traditional
17.	Accidental bank transfer scam	Scammers claim accidental transfers to demand refunds.	Traditional
18.	Fraudulent recovery services	Scams claiming to recover lost money.	Traditional
19.	Telegram OTP theft	Stealing Telegram OTP codes to gain access.	Traditional
20.	Fake money loss call scams like FlashAI	Spreading false news about losing money through calls.	AI Technology
21.	Facebook recovery service scam	Scams claiming to recover lost Facebook accounts.	Traditional
22.	Romance and financial investment scams	Luring victims with love, fake investments, parcel deliveries, or lottery winnings.	Traditional
23.	Phishing links and fake ads on Facebook	Fraudulent links and deceptive advertisements.	AI Technology
24.	Lottery number scam	Providing fake lottery number predictions.	Traditional

Source: MIC and author's compilation

Annex B. Target Groups and Types of Scams Associated With Each Group

Target Group	Types of Scams	AI Technology or Traditional
Elderly	1. Travel package scams with "cheap combos."	Traditional
	2. Scams involving Deepfake video calls.	AI Technology
	3. "SIM lock" scams due to incomplete registration of phone numbers.	Traditional
	4. Impersonation for successful money transfers.	Traditional
	5. Fake messages impersonating government, enterprises, or organizations (e.g., social insurance, banks).	Traditional
	6. Fake brand name promotional messages.	Traditional
	7. Impersonation of police, investigators, courts, via fraudulent phone calls.	Traditional
	8. Scams involving low-quality goods on e-commerce platforms.	Traditional
	9. Stealing personal information from ID cards for fraudulent activities.	Traditional
	10. Fake accidental transfers to bank accounts.	Traditional
	11. Service scams targeting Facebook account recovery.	Traditional
	12. Emotional manipulation, investments, or fraudulent packages.	Traditional
	13. Phishing links via fake advertisements on Facebook.	Traditional
	14. Scams involving betting or gambling.	Traditional
	15. Spreading fake news about losing money.	Traditional
Children (under age 18)	1. Scams involving Deepfake video calls.	AI Technology
	2. Scams with emotional manipulation or sharing sensitive images.	Traditional
	3. Facebook account recovery service scams.	Traditional
Students/ Youth	1. Travel package scams with "cheap combos."	Traditional
	2. Scams involving Deepfake video calls.	AI Technology
	3. "SIM lock" scams due to incomplete registration of phone numbers.	Traditional
	4. Fraudulent gambling apps, betting, or black-market links.	Traditional
	5. Fake brand name promotional messages.	Traditional
	6. Financial fraud or fake investment scams.	Traditional
	7. Fake online collaborator recruitment.	Traditional
	8. Impersonation of police, investigators, courts, via fraudulent phone calls.	Traditional
	9. Scams involving low-quality goods on e-commerce platforms.	Traditional
	10. Stealing personal information from ID cards for fraudulent activities.	Traditional
	11. Fake accidental transfers to bank accounts.	Traditional
	12. Service scams targeting Facebook account recovery.	Traditional
	13. Emotional manipulation, investments, or fraudulent packages.	Traditional

Source: MIC and author's compilation

REFERENCES

ASEAN. (2021). *ASEAN Digital Economy Framework Agreement.*

Chongluadao. (2025). *Available at:* https://chongluadao.vn/thong-ke?type=blacklist

GASA and Gogolook (2023). *Asia Scam report 2023*. Available at : https://hpt.vn/Uploads/File/2023/Bao-cao-lua-dao-Chau-A-2023.pdf

Global Anti-Scam Alliance (GASA) and Chongluadao.vn. (2023). *The State of Scams In Vietnam. 2023*. Available at: https://thesaigontimes.vn/wp-content/uploads/2024/01/State-of-Scam-Report-2023-Vietnam.pdf

National Cyber Security Center (NCSC). (2023). *Cybersecurity report 2023*

Vietnam Ministry of Information and Communications. (2022). *Cybersecurity Report 2022.*

Vietnamnet. (2025). Dismantling a cross border scam phone call ring, appropriating nearly 1000 billion VND. Available at: https://vietnamnet.vn/triet-pha-duong-day-goi-dien-lua-dao-xuyen-bien-gioi-chiem-doat-1-000-ty-dong-2366553.html

World Economic Forum. (2023). *Global Cybersecurity Outlook 2023*

Safer Internet Lab

saferinternetlab.org

Jl. Tanah Abang III no 23-27
Gambir, Jakarta Pusat. 10160

Find Us On

CSIS Indonesia | Safer Internet Lab