Safer Internet Lab

Research report

Online Fraud and Scams in South Korea

Safer Internet Lab

ONLINE FRAUD AND SCAMS IN SOUTH KOREA

Safer Internet Lab

A Research Report by Safer Internet Lab

The views expressed here are solely those of the author(s) and do not represent an official position of SAIL, CSIS, Google, or any other organization. Please contact the author(s) directly with any comments or questions.

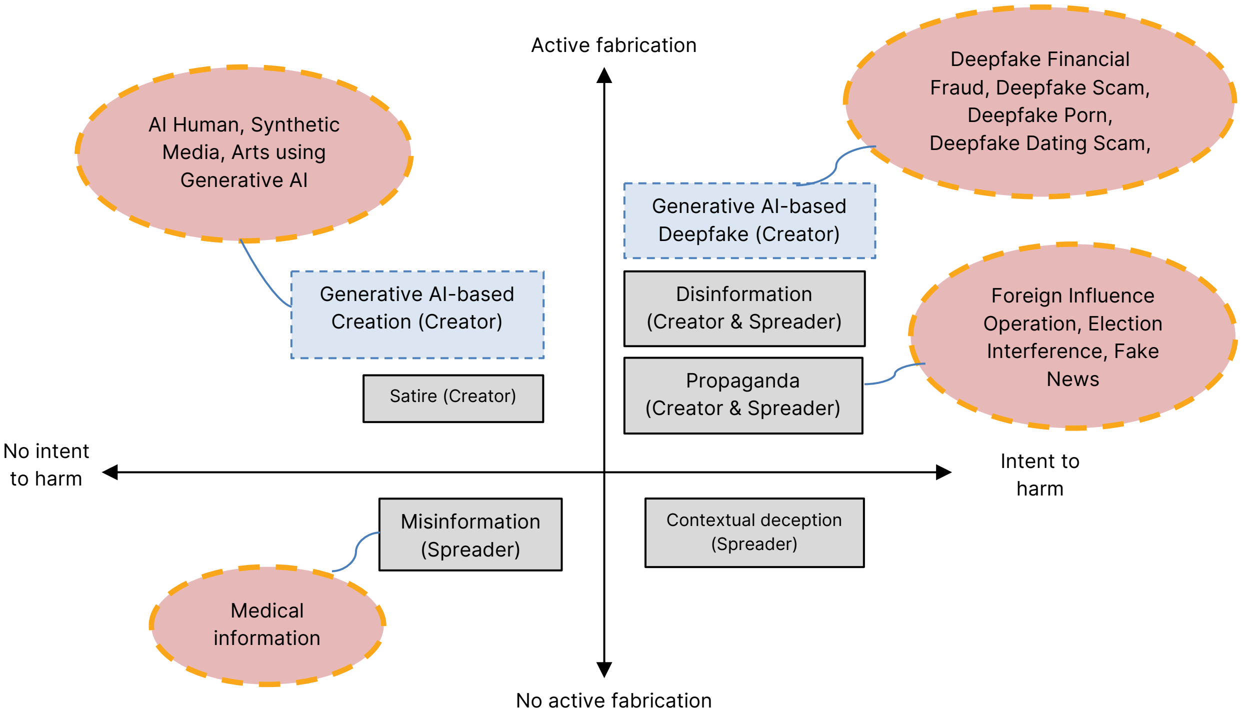
Online Fraud and Scams in South Korea

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


Active fabrication
AI Human, Synthetic Media, Arts using Generative AI
Deepfake Financial Fraud, Deepfake Scam, Deepfake Porn, Deepfake Dating Scam,
Generative AI-based Deepfake (Creator)
Generative AI-based Creation (Creator)
Disinformation (Creator & Spreader)
Foreign Influence Operation, Election Interference, Fake News
Satire (Creator)
Propaganda (Creator & Spreader)
No intent to harm
Intent to harm
Misinformation (Spreader)
Contextual deception (Spreader)
Medical information
No active fabrication

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

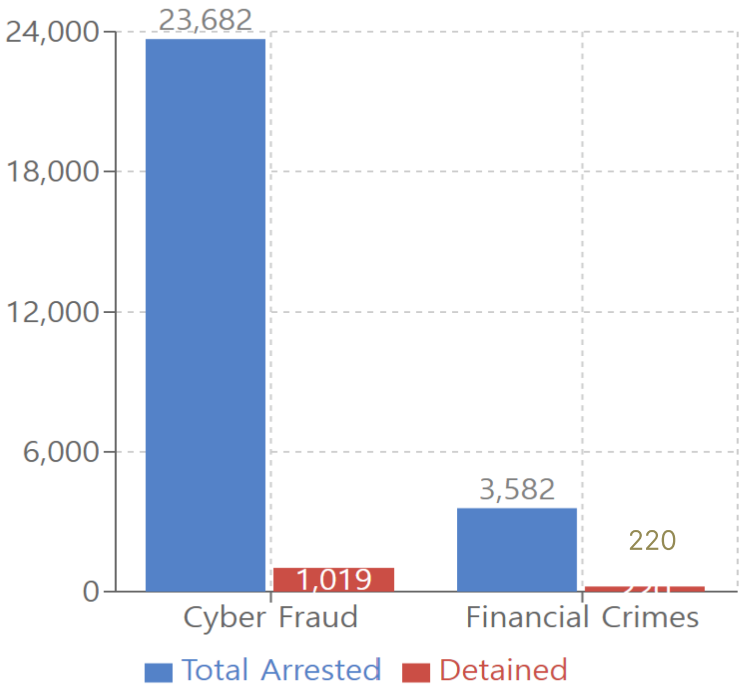
South Korea has experienced a rise in sophisticated financial and online fraud cases leveraging generative AI and deepfake technology. The increasing use of generative AI threatens traditional policy measures and prevention approaches. Online scams evolved to adopt generative AI technology and created 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.

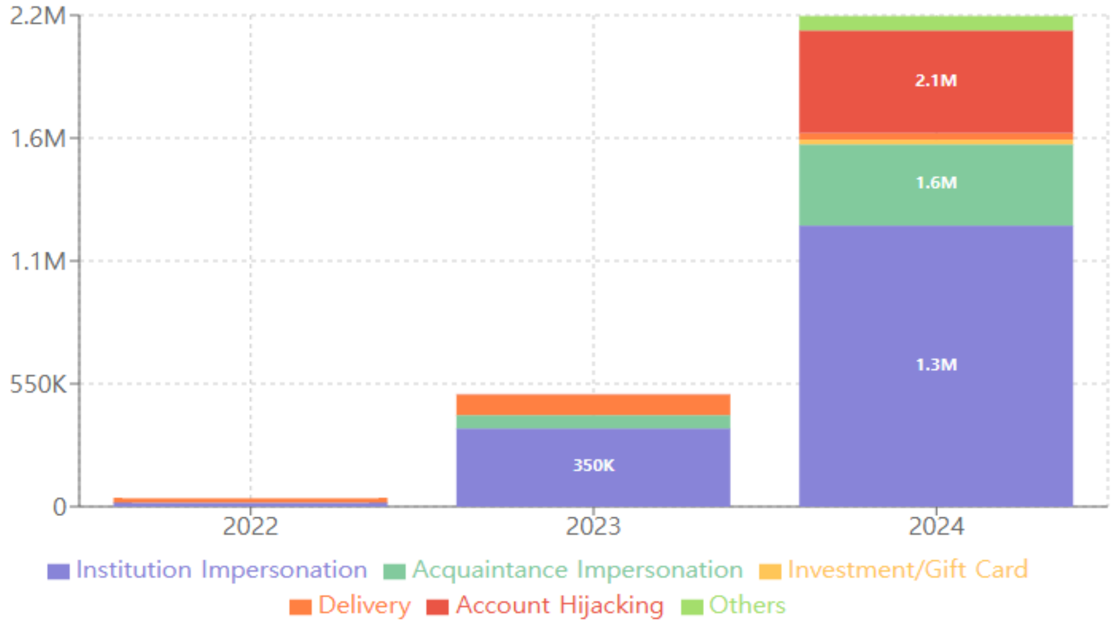
Figure 8.2 Cyber fraud and financial crime arrests in 2023



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.

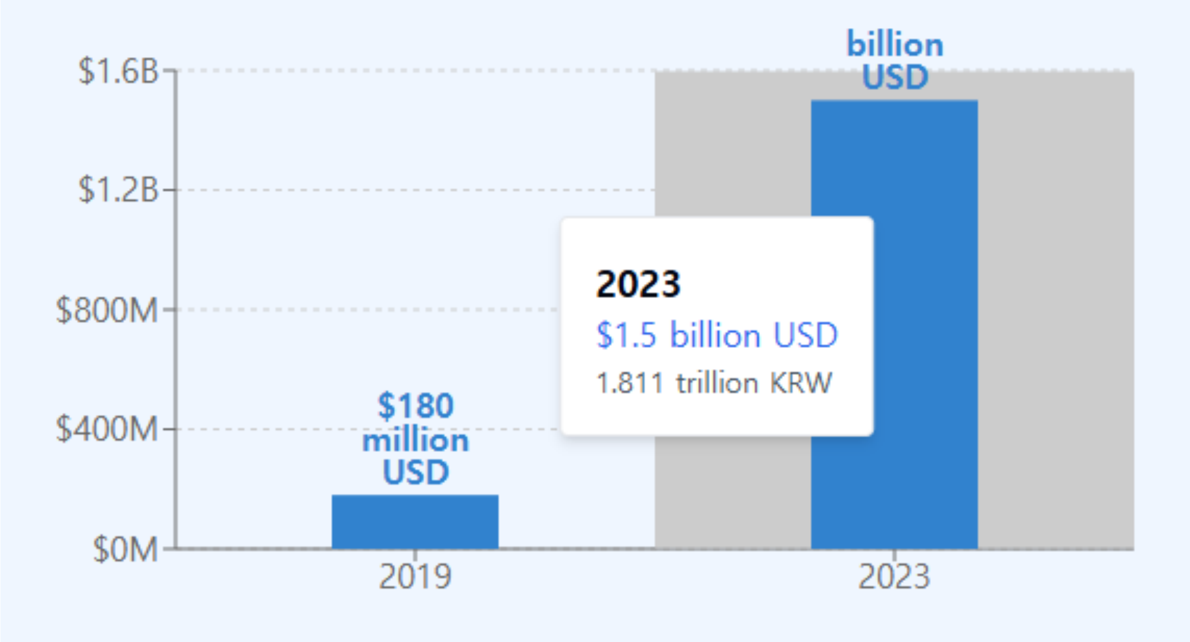
These patterns and trends highlight a critical point in South Korea's battle against cyber fraud and the emerging threats posed by the misuse of AI technology. As generative AI technologies become more accessible and sophisticated, the line between traditional cyber scams and AI-enhanced fraud continues 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.

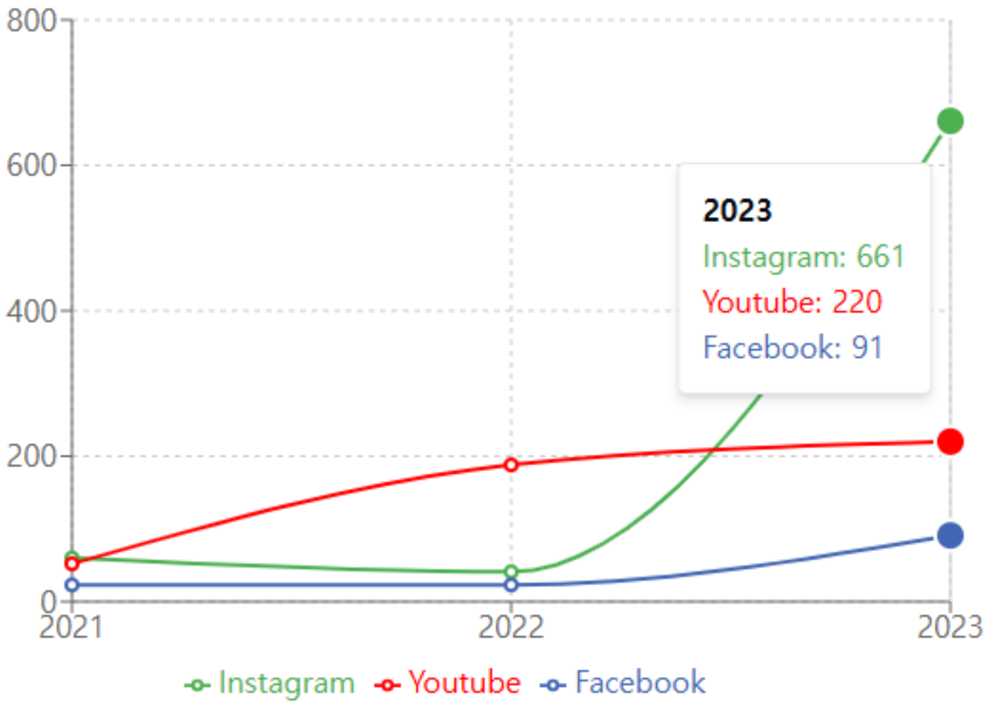
In this way, online scams around the world definitely push for security-first development. Financial sectors will increase their cyber security efforts, but malicious actors almost always find a way to circumvent or attack the secured network. With the development of AI-technology, it is getting easier 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

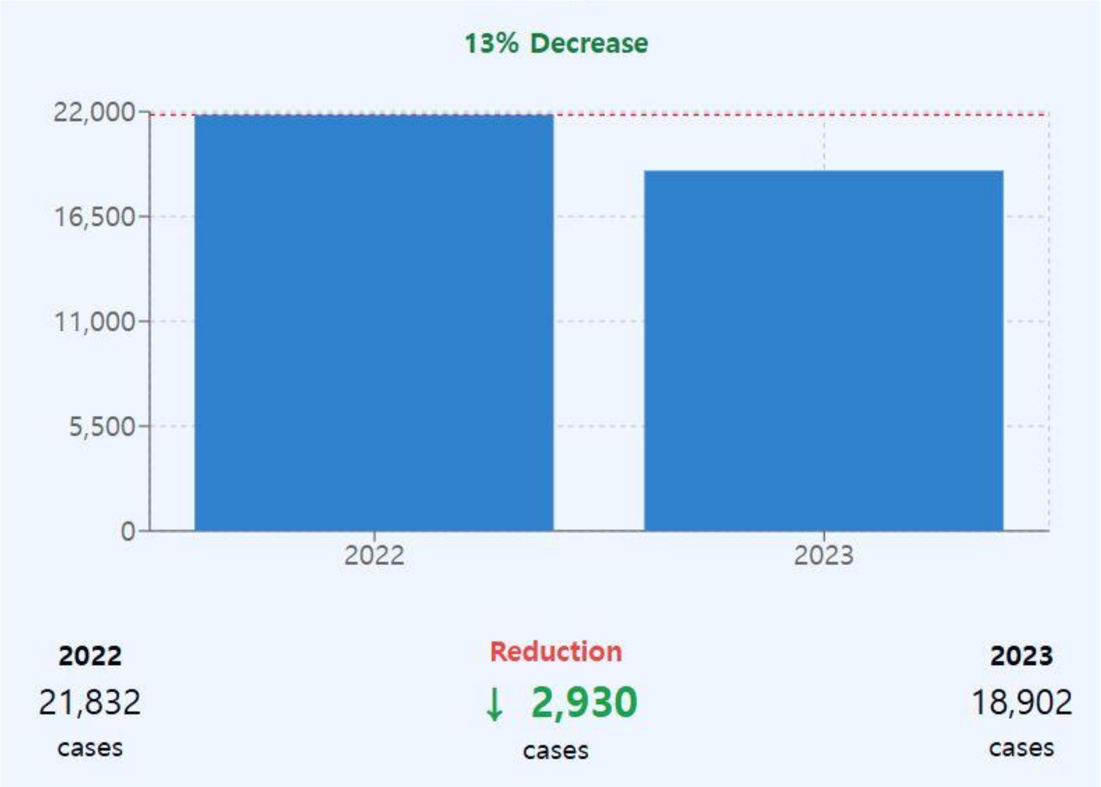
There are mainly two types of proactive or preventive policy responses offered by the South Korean government: 1) technical response and reporting system, and 2) public awareness and education.

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), “ㅇㅇㅇ인데요” 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

Safer Internet Lab

saferinternetlab.org

Jl. Tanah Abang III no 23-27
Gambir, Jakarta Pusat. 10160

Find Us On

CSIS Indonesia | Safer Internet Lab