Л.Н. Гумилев атындағы Еуразия ұлттық университетінің ХАБАРШЫСЫ.
ISSN: 2616-6844. eISSN: 2663-1318

ҚҰҚЫҚ СЕРИЯСЫ/ LAW SERIES/ СЕРИЯ ПРАВО

Қылмыстық құқық. Қылмыстық процесс. Криминалистика. Криминология / Criminal law. Criminal process. Criminalistics. Criminology / Уголовное право. Уголовный процесс. Криминалистика. Криминология

IRSTI 10.77.51
https://doi.org/10.32523/2616-6844-2025-152-3-188-211
Scientific article

Deepfake technologies and social engineering in online fraud forms, mechanisms, and legal challenges

A.B. Smanova*[1], A.Zh. Muratova[2], Sh.R. Zhumagulova[3]

[1]al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Mukhametzhan Tynyshbayev ALT University, Almaty, Kazakhstan
[3]Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan

(e-mail: [1]akmaralbahtyar@gmail.com, [2]kkaebsong98@mail.ru, [3]jumagulova_sholpan@mail.ru)

Abstract: The digital world has undergone fundamental changes through generative neural networks that are advancing at a fast pace in artificial intelligence development. Deepfake technologies enable users to generate realistic audio and video content that duplicates actual recordings. The combination of these tools with social engineering techniques makes fraudulent schemes more believable which results in financial losses and increased cybercrime activities. Multiple jurisdictions, including Kazakhstan, face challenges in creating effective legal frameworks to address these emerging technological developments.

The research investigates deepfake technology and social engineering methods used in online scams while identifying legal obstacles to prevention and proposing solutions for the improvement of legislation and institutional practices.

The scientific value and practical relevance of the work lie in clarifying the interaction between psychological manipulation and generative AI, as well as in providing proposals for shaping a national strategy to counter cyber threats. The methodology relies on a comprehensive analytical approach that includes the study of documented cases of synthetic media misuse, comparative analysis of international and national legislation, and assessment of preventive strategies.

The findings show that deepfakes are widely applied in executive imper-sonation scams, circumvention of voice biometric systems, investment fraud using fabricated images of public figures, and privacy violations such as extortion. The key legal challenges include difficulties in crime qualification, authentication of digital evidence, cross-border dimensions of cyberattacks, and allocation of liability between perpetrators and online platforms.

The study concludes that an integrated approach is required, combining technological tools for the detection and labeling of synthetic media, legal reforms to strengthen criminal and procedural law, harmonization of international norms, and educational measures to improve digital literacy. This research contributes theoretical and practical foundations for a systemic response to cybercrime involving deep synthesis technologies.

Keywords: deepfake technologies, social engineering, online fraud, artificial intelligence, legal challenges, cybercrime.

Received: 16.09.2025. Accepted: 26.09.2025. Available online: 30.09.2025

***corresponding author**

Introduction

The rapid advancement of artificial intelligence technologies, particularly generative neural networks, has introduced a qualitatively new level of threats in the digital environment. One of the most prominent manifestations of this trend is deepfake technology, which enables the creation of synthetic audio and video content that is virtually indistinguishable from authentic material. Initially perceived as an innovation in the entertainment industry, such tools have quickly acquired widespread criminal applications, highlighting the growing criminogenic potential of digital innovation. Fraudsters leveraging AI-generated deepfakes now represent a rising cybersecurity threat, underscoring the urgent need not only to harness AI's benefits but also to mitigate its weaponization.

AI-driven systems are increasingly used to fabricate information across multiple formats – text, audio, images, and video – making it ever more difficult to distinguish authentic content from falsified data. The deployment of such systems by parties in armed conflicts to amplify propaganda, manipulate public opinion, and influence decision-making may carry serious consequences [1].

This dynamic is confirmed by national statistics: according to the National Computer Incident Response Center (KZ-CERT), the number of reported cybercrimes in Kazakhstan has doubled over the past two years – from 34,500 incidents in 2023 (a 107% increase from 2022) to 68,100 cases in 2024 (a further 97% rise) [2]. The global picture reveals a similar trajectory. According to Regula's The Deepfake Trends 2024 report [3], the average financial loss from synthetic media fraud now reaches USD 450,000 for most organizations, exceeding USD 603,000 in the financial services sector. Strikingly, 92% of surveyed companies reported losses of up to USD 450,000, while 10% faced damages exceeding USD 1 million. For comparison, the 2022 average was approximately USD 230,000–nearly half of today's level. Fintech companies are particularly vulnerable, with average losses of USD 637,000 compared to USD 570,000 in traditional banking. Geographically, the heaviest losses are reported in Mexico (USD 627,000), Singapore (USD 577,000), and the United States (USD 438,000).

An alarming gap has emerged between organizational confidence and preparedness: although 56% of companies report high confidence in their ability to detect deepfakes, only 6% have successfully avoided financial damage. This imbalance illustrates the business sector's insufficient readiness for increasingly sophisticated attacks, especially within financial services. The convergence of deepfake content with social engineering poses a particular danger. Exploiting trust and psychological manipulation, such tactics in a globalized information space have become powerful tools of cybercrime, dramatically amplifying the persuasiveness of fraudulent schemes. Reports from law enforcement and international research centers emphasize that social engineering remains one of the key enablers of cyberattacks, while the integration of synthetic media technologies significantly enhances their effectiveness, generating new risks for financial institutions, government bodies, and individuals alike.

International practice confirms the scale of the problem: damages from deepfake-based crimes are already measured in tens of millions of dollars. One illustrative case occurred in January 2024, when criminals defrauded the engineering company Arup of USD 25.5 million through a sophisticated deepfake operation [4]. A finance officer in Hong Kong, convinced he was speaking with his UK-based CFO and several colleagues during a video call, authorized 15 transfers amounting to USD 25.5 million. Weeks later, it was revealed that all participants except the victim were AI-generated deepfakes. This incident highlights more than just fraudulent ingenuity; it signals a fundamental disruption of the trust infrastructure underpinning modern business. It further demonstrates why organizations seeking to benefit from AI must also defend against its misuse. Similar cases have been reported worldwide, where employees approved major transactions after being deceived by fabricated video conferences with “executives.” These developments confirm that deepfake fraud has transcended local contexts, evolving into a global challenge with implications spanning economics, politics, and national security.

The urgency of this issue is compounded by the inadequacy of existing legal frameworks. National criminal codes typically address conventional fraud and identity theft, but rarely account for crimes enabled by synthetic media. As a result, law enforcement faces difficulties in classification, evidentiary practice, and cross-border prosecution. Collectively, these gaps underscore the need for a comprehensive analysis of the forms, mechanisms, and legal challenges associated with the use of deepfake technologies and social engineering in online fraud–an endeavor that is both scientifically and practically significant.

Recent studies in information security reveal a steady growth in crimes driven by generative AI technologies. The pace of this evolution is striking: cases of deepfake fraud in North America increased by 1,740% between 2022 and 2023, with financial losses surpassing USD 200 million in the first quarter of 2025 alone. The democratization of deepfake tools has lowered entry barriers for fraudsters: voice cloning requires only 20–30 seconds of audio, while convincing video deepfakes can be produced within 45 minutes using freely available software [4]. These tools allow criminals to generate synthetic images, audio, and video with high fidelity to specific individuals, creating the illusion of genuine interaction and enabling unauthorized access to financial and confidential assets.

Corporate fraud has emerged as one of the most prevalent applications. Criminals imitate executives’ voices to issue false instructions for fund transfers or access to strategic data. In some cases, employees have participated in video conferences populated by multiple AI-generated “colleagues,” resulting in severe financial losses. Such incidents demonstrate how the combination of social engineering and deepfakes enables high-level deception that is nearly impossible to detect in real time [5].

Beyond the corporate sphere, synthetic media technologies are widely exploited in online marketing and investment scams. Fabricated videos depicting public figures, entrepreneurs, and politicians “endorsing” investment platforms or cryptocurrency schemes foster trust among potential victims and accelerate the spread of fraud.

Another alarming dimension involves privacy violations and reputational harm. Deepfakes are increasingly used to produce intimate images and videos without consent, often in connection with extortion and blackmail. These practices not only inflict severe psychological harm on victims but also fuel the emergence of new forms of cyber-violence.

Furthermore, deepfake technologies are actively applied in identity fraud. Synthetic photographs and video streams can deceive biometric authentication systems employed by banks and government agencies [6]. This development undermines confidence in modern cybersecurity tools and introduces additional risks for financial and national security.

Deepfakes used in social engineering attacks create multiple security risks, including financial losses, damage to reputation, and physical harm to individuals, as illustrated by real-world examples. The advancement of generative AI technology will lead to an increase in both the number and sophistication of such crimes. The growing complexity of deepfake-related crimes requires scholars, legal experts and international organizations to dedicate increased attention to this issue.

The research investigates deepfake and social engineering methods used in online scams and identifies essential legal barriers for their prevention. Specifically, it will: (1) identify the main forms and scenarios of synthetic media in criminal practice; (2) study how generative AI supports psychological manipulation in fraud; (3) compare different national and international approaches to regulation and prevention; (4) reveal the main legal and procedural problems in investigating and prosecuting such crimes; and (5) offer recommendations for improving laws and institutions to provide more effective protection against fraud driven by AI technologies.

Methodology

The object of this study is deepfake technologies and social engineering in online fraud. The material includes academic publications, legal acts, statistical reports, and case descriptions, combining both qualitative and quantitative data. The research applies system analysis, comparative analysis, content analysis, case study, classification and typologization, as well as synthesis and generalization. These methods made it possible to view the problem as an interplay of technological, legal, and social factors, compare international approaches, systematize fraud forms, and formulate comprehensive conclusions and recommendations.

Findings/Discussion

One of the most dangerous manifestations of synthetic media in online fraud involves financial manipulations carried out through the impersonation of executives or other officials during video conferences. This method relies on creating highly convincing visual and auditory forgeries that reproduce the appearance, facial expressions, and voice of an authorized individual capable of issuing instructions for financial transfers or granting access to confidential data [7; 8].

Recent practice demonstrates that such attacks are becoming increasingly systematic. In several cases, employees of corporate finance departments received instructions from fabricated “executives” during video calls whose visual and audio quality raised no suspicion. One of the most widely reported incidents involved the multinational corporation Arup, where fraudsters simultaneously imitated several senior managers in a deepfake video conference [9].

This type of fraud poses a particular threat to corporate security because it exploits not only the technological sophistication of deepfakes but also internal organizational vulnerabilities. Criminals leverage employees’ trust in authority, corporate cultures that prioritize rapid compliance with executive directives, and time pressure during business communication. These conditions create an environment in which even experienced professionals may fail to question the authenticity of incoming instructions.

Deepfake-enabled video call fraud is further characterized by high latency and cross-border complexity. Stolen funds are typically transferred through international banking networks and cryptocurrency platforms, making tracking and recovery extremely difficult. Investigations face additional challenges in establishing perpetrator identities and in the admissibility of digital evidence based on synthetic media analysis. Given the scale and potential consequences of these crimes, prevention requires both technical measures (deepfake detection and identity verification tools for online communications) and organizational safeguards (mandatory multi-level approval of financial transactions). Without such a comprehensive approach, corporate financial stability remains at significant risk.

A particularly vulnerable area of financial security involves the use of voice-cloning technologies to circumvent client identification procedures. Contemporary machine learning algorithms can reproduce a person's voice with remarkable precision, replicating timbre and intonational patterns. In many cases, fraudsters need only a few minutes of original audio, obtained through open sources, phone conversations, or data leaks, to generate convincing forgeries.

The risks are especially acute in banking, where voice-based biometric authentication is increasingly deployed. Successful cloning allows criminals to pass verification in call centers, initiate transactions, or alter account settings. In documented cases, attackers executed transfers and contracted loans using only AI-generated voice commands[10].

The threat escalates when combined with social engineering techniques. Fraudsters construct scenarios emphasizing urgency and pressure, which reduces the likelihood of additional checks. The use of telephone channels further minimizes the chance of visual verification, leaving voice as the sole marker of authenticity.

Investigations into such crimes face significant evidentiary obstacles. Forensic identification of AI-generated voices requires advanced acoustic analysis methods, yet the increasing precision of generative models makes detection extremely challenging [11]. Additional difficulties arise from the transnational nature of these attacks: access to biometric voice data may originate outside the victim state's jurisdiction, limiting the capacity of law enforcement.

As voice-cloning technologies become more accessible, the need for comprehensive countermeasures intensifies. Protection of organizations and their clients from this threat requires a harmonious blend of legal frameworks and technological solutions.

Implementing effective security measures, such as combined biometric and behavioral verification methods, necessitates robust international and national regulatory frameworks. However, the digital environment now faces a new, distinct threat: investment scams that use deepfake content featuring well-known public figures. The creation of realistic audio and video segments by criminals shows famous people promoting financial products and cryptocurrency schemes, which appear genuine to unsuspecting victims. The deceptive appearance of these fake materials makes victims less cautious while allowing scammers to achieve higher success rates in their fraudulent activities.

The research investigates deepfake technology alongside social engineering tactics used in online scams while identifying key legal barriers that prevent their detection.

Cybersecurity research highlights the widespread nature of such crimes, particularly on social networks and video-hosting platforms, where recommendation algorithms accelerate the dissemination of fraudulent clips. These materials often contain links to fictitious investment platforms promising guaranteed returns or access to “exclusive” projects. The deception relies on public trust in recognizable figures and on the authority principle, thereby amplifying the social engineering effect [12].

Such crimes are marked by high latency: victims often realize the fraud only after transferring funds and refrain from reporting incidents, believing recovery prospects to be minimal. Investigations are complicated by offshore jurisdictions and anonymous cryptocurrency transactions, which hinder the identification of perpetrators and restitution of losses [13].

From a legal standpoint, these actions may be classified as fraud involving the unlawful use of personal data and likeness. Yet, current regulations frequently fail to address the specifics of digital technologies that enable mass distribution of synthetic content, creating significant obstacles to effective enforcement [14]. Considering the growing scope of the problem, the development of international standards for regulating generative AI in media has become essential. Mandatory requirements for platforms–such as labeling synthetic content and swiftly removing fraudulent materials – could reduce the prevalence of such schemes and strengthen digital trust.

Among the most socially harmful applications of synthetic media is the creation and dissemination of pornographic deepfake content, often accompanied by blackmail (sextortion). This category of crimes is highly latent and exerts profound psychological pressure on victims. Generative algorithms enable face-swapping in images and videos with a level of realism that makes falsifications nearly indistinguishable from authentic materials. Criminals exploit this for reputational harm, coercion, and extortion of money or other benefits.

The threat is particularly acute when targeting women and minors. Digital criminology research records a growing number of deepfake pornography incidents affecting public figures, journalists, politicians, entertainers, and private individuals lacking resources to defend their rights effectively [15]. The rapid spread of such content via social networks and messaging platforms leads to long-lasting reputational and psychological damage.

Sextortion in the digital environment often assumes a complex form. Offenders not only disseminate forged images or threaten their publication but also combine these tactics with social engineering, coercing victims into actions that serve criminal interests. The widespread availability of generative tools, which require no advanced technical expertise, further fuels the expansion of this form of cybercrime [16].

From a legal standpoint, such actions are usually associated with breaches of privacy, the illegal distribution of pornographic content, and various forms of extortion. Yet, the current legal norms often turn out to be insufficient, since they do not take into account the distinctive features of synthetic media. Issues such as whether deepfake materials can be accepted as valid evidence and the difficulties connected with their cross-border circulation still demand more thorough legal analysis.

Academics, together with practitioners, stress the need to develop specific legal safeguards that protect people from sexualized digital violence. The proposed solutions include making it illegal to create and distribute unauthorized pornographic deepfakes and developing efficient content removal systems and international cooperation for digital forensic work. The combination of these strategies serves as fundamental measures to fight sextortion threats while safeguarding personal rights in the fast-changing landscape of generative AI technology.

Modern deepfake-enabled attacks depend on recent developments in generative modeling, which create synthetic content that appears highly realistic. The two primary architectures used for content generation are generative adversarial networks (GANs) and diffusion models, which produce realistic synthetic videos and images with authentic facial movements and natural speech patterns. The evolution of deep learning algorithms has made it possible to create synthetic avatars of specific individuals from relatively short datasets, reproducing characteristic micro-expressions and speech patterns.

Similar progress has been achieved in audio signal processing. Neural speech synthesis and real-time voice-conversion systems now allow not only precise timbre cloning but also adaptation to emotional states and intonations. Integrated into video conferencing platforms, these tools create the illusion of authentic communication, depriving interlocutors of visual or acoustic cues that might otherwise indicate manipulation.

A key factor amplifying the risk of deepfake-based attacks is the emergence of real-time generation technologies. Today, both audio and video can be produced online with minimal latency, enabling criminals to impersonate specific individuals during calls without suspicious pauses [17]. The availability of open-source solutions and commercial services offering real-time synthetic media generation further magnifies the threat. By lowering technical entry barriers, these tools extend beyond organized criminal groups to individual actors with only basic IT skills. Consequently, deepfakes are no longer a niche technological experiment but a mass instrument of cybercrime, threatening financial stability, information security, and public trust in digital communications.

The effectiveness of such schemes is closely tied to social engineering strategies designed to manipulate victims' cognitive and emotional states. Psychological mechanisms of trust and authority play a decisive role in decision-making under informational uncertainty. When the visual and vocal appearance of a source aligns with expectations, the likelihood of critical evaluation of the message declines sharply [18].

Authority thus becomes a fundamental tool of influence in deepfake-enabled attacks. The use of executive or official appearance and voice in professional or financial settings makes people more susceptible to deception. The reproduction of familiar behavioral patterns, including facial expressions, nonverbal signals, and intonations, creates a sense of authenticity that strengthens trust.

The establishment of a sense of urgency stands as a vital component in these situations. The fraudsters create situations that force victims to act quickly because they present themselves as urgent matters that need immediate resolution. The sense of urgency makes people less likely to verify information while making them more prone to act on impulse. The psychological pressure may take the form of dissatisfaction statements or professional obligation reminders or threats about adverse outcomes when someone refuses.

People who understand deepfake technology might still miss warning signs when authority figures create a sense of urgency. This underscores the necessity of a comprehensive preventive approach that includes not only the advancement of detection technologies but also training programs to strengthen awareness of manipulative psychological techniques [19].

The integration of generative AI technologies with social engineering produces a synergistic effect that significantly enhances the success of criminal schemes. While deepfakes provide visual and acoustic realism, social manipulation exploits cognitive vulnerabilities, reducing critical analysis. The result is a situation where technological and psychological elements reinforce one another, forming a resilient deception model [18].

The technical component ensures consistency and plausibility in synthetic personas, synchronizing facial expressions, voice, and nonverbal signals while eliminating traditional indicators of deceit. Simultaneously, psychological pressure – through urgency or appeals to authority – minimizes the likelihood that recipients will attempt independent verification. The convergence of these factors yields a comprehensive attack model in which each element strengthens the other. High media realism fosters trust, while manipulative strategies suppress rational assessment, leaving victims unprepared to respond effectively – even when they are aware of deepfake risks.

This synergy amplifies the transnational danger of such schemes, as effective counteraction requires simultaneous advancement of both detection technologies and resilience-building programs for users. Without an integrated strategy, cybercrimes that combine deepfake modeling with social engineering remain especially resistant to conventional protective measures [20].

The use of deepfake technologies and social engineering in online fraud presents legal systems with complex and unprecedented challenges. Traditional mechanisms of criminal law, designed to combat classic forms of deception and identity theft, prove inadequate in the rapidly evolving context of generative technologies. The central difficulty lies in the fact that modern legal frameworks do not always account for crimes involving synthetic media, where the boundary between authentic and fabricated information is increasingly blurred [14].

A major concern is the legal qualification of crimes involving deepfakes, which is marked by uncertainty and the absence of unified approaches in international practice. Substantial difficulties also arise in evidentiary proceedings, as courts face challenges in assessing the admissibility of digital evidence when traditional standards of authenticity verification are ill-suited to synthetic content. The transnational nature of online crimes further aggravates the situation, with offenders exploiting differences in national jurisdictions and gaps in international cooperation to evade accountability.

Another unresolved issue is the allocation of responsibility between direct perpetrators and the platforms that enable the creation and dissemination of deepfake content. Ongoing debates on the extent of liability for service providers and platform operators illustrate the pressing need for new legal standards that balance the protection of individual rights with the promotion of technological innovation [21]. Collectively, these factors highlight the necessity of a systematic reconsideration of legal challenges associated with synthetic media in fraudulent schemes. Such an analysis must address issues of crime qualification, evidentiary standards, cross-border dimensions, and responsibility allocation within the digital ecosystem.

The classification of deepfake-related offenses stands as a major challenge for modern criminal law doctrine. The existence of sophisticated audio and video forgery algorithms enables criminals to execute schemes that do not fully align with present criminal laws. The lack of dedicated laws forces law enforcement to apply conventional categories, which leads to irregular practices and diminished operational success.

The most prevalent legal classification for these offenses is fraud because deception and trust exploitation for monetary gain are core elements of such crimes. While the use of synthetic media fits within the classical framework of fraud, the technological specificity of deepfakes requires clearer delineation of liability. The difficulty arises from the fact that the deception is not based on direct false statements but on the use of technical instruments that create an illusion of authenticity.

Identity theft constitutes another important dimension. The falsification of facial images, voice, or other identifying features via generative technologies may qualify as unlawful appropriation or use of another person's identifying data. In some legal systems, such actions are classified as separate criminal offenses. At the same time, many jurisdictions still debate whether data generated synthetically can fall under the category of personal data, and no clear consensus has been reached.

Of particular concern is the misuse of biometric identifiers. Current data protection frameworks generally regard voice, facial images, and other unique biological markers as sensitive information that requires stronger safeguards. When deepfake materials are created or distributed without a person's consent, this undermines both individual autonomy and the right to privacy, and thus ought to be recognized as an offense in its own right. The lack of specific rules about falsified biometric data treatment in legal orders generates uncertainties during court cases.

media as their instrument of operation. The recognition of deepfakes as an independent criminal tool would. The current situation demonstrates an urgent requirement for criminal law to create specific provisions that handle crimes that use synthetic media, thereby ensuring better protection and filling existing legal gaps while creating standardized procedures for dealing with fraud-related and identity theft and biometric data misuse offenses.

The main challenge in deepfake crime investigations stems from the difficulty of obtaining reliable evidence. The current criminal procedure system depends on evidence that can be verified and proven to be reliable. The foundation of traditional evidence verification has been compromised by synthetic media, according to recent research [22], because it produces convincing artificial content that experts struggle to detect.

The main difficulty in digital evidence authentication remains a critical issue. Given that advanced generative algorithms now achieve high levels of synchronization in speech, facial expression, and motion, experts struggle to distinguish between authentic and artificial content. This threatens the principle of admissibility, as courts cannot rely on evidence whose validity cannot be confirmed by objective and reproducible methods [23].

Additional difficulties stem from the absence of standardized forensic procedures for deepfake analysis. Different jurisdictions apply divergent methodologies, producing inconsistent expert conclusions and undermining judicial trust. In some legal systems, there is no explicit regulation of procedures for identifying synthetic media, enabling defense parties to challenge the legality and reliability of forensic findings.

The evidentiary value of digital materials is further complicated in cross-border investigations [24]. The process of electronic evidence transfer between states encounters three main obstacles, which stem from inconsistent data formats and conflicting legal frameworks for data preservation and authentication and insufficient international agreements about evidence acceptance standards.

A solution to these problems requires a unified approach that develops technical standards for deepfake detection and establishes standardized forensic methods for synthetic media analysis and creates international rules for digital evidence sharing. The absence of protective measures makes it challenging to achieve technological advancement alongside fair trial protections.

Deepfake and social engineering crimes share the distinctive feature of operating across international borders. Offenders can use digital networks to perform operations from their home country while inflicting damage on victims located elsewhere. The lack of geographical boundaries in digital networks creates challenges for law enforcement agencies to track perpetrators and obtain valid evidence and establish legal responsibility.

The main challenge arises because different countries have separate laws regarding deepfake activities. The lack of equivalent legal frameworks between nations makes it impossible to prosecute deepfake offenses through existing laws that apply in some jurisdictions. The absence of matching laws between countries makes it difficult for countries to work together on extradition and the prosecution of criminal cases.

The situation becomes more complex because of how perpetrators use anonymization tools and decentralized financial systems. The combination of cryptocurrency systems with online privacy tools makes it harder for law enforcement to identify suspects and execute court orders. States that work together on investigations face extended periods of time before they can obtain justice because they lack sufficient power to capture perpetrators.

International legal standards that focus on deepfake criminal activity need to be developed through coordinated efforts between nations. The international community can develop an effective digital crime response system through standardized national laws and enhanced information sharing between states for better protection in globalized environments.

The main point of contention in this area concerns which party bears responsibility for digital content distribution between platform operators and content creators. The distribution of synthetic content occurs mainly through platforms, which positions them as essential entities for maintaining cybersecurity and safeguarding user rights. The European Digital Services Act (DSA) implements due diligence requirements for platforms to stop unlawful content distribution, including deepfakes, while establishing labeling systems for better transparency.

In common law jurisdictions, the emphasis is on limiting the immunity of online intermediaries, which allows platforms to be held liable for inaction in the face of clear user rights violations. By contrast, a technology-neutral approach, more liberal in nature, releases platforms from liability for user-generated content, thus fostering innovation but weakening responsiveness to abuses of deepfake technologies [26].

The global reach of digital platforms creates an even greater challenge. Even where national laws mandate moderation, enforcement is difficult when platforms are registered outside the relevant jurisdiction. This underscores the importance of harmonizing international norms and developing cross-border monitoring and cooperation mechanisms [27].

Balancing rights protection with technological progress is therefore critical. The implementation of overly strict rules creates barriers to innovation but insufficient oversight leads to abuse and damages public trust. Legal experts propose implementing a multi-layered regulatory framework that merges industry self-regulation with government oversight and international standards to create a flexible yet standardized system.

Deepfake misuse has become a worldwide concern, which prompted international bodies to create new legal frameworks to handle this problem at both national and international levels. Countries have adopted two different methods to handle synthetic media misuse through either legal reinterpretation or the creation of new laws that restrict its harmful applications. Different jurisdictions implement distinct approaches because their legal systems and technological resources vary substantially.

Two primary international approaches have emerged during current global dialogues. One approach advocates for the adoption of dedicated legal instruments regulating artificial intelligence, including rules on transparency, labeling, and accountability in the deployment of deepfake technologies. For example, the EU Artificial Intelligence Act (AI Act) requires that synthetic content be labeled in a machine-readable format, as well as visibly marked to indicate its artificial origin [28]. The second strategy involves integrating provisions on synthetic media into existing frameworks on data protection, anti-fraud measures, and cybercrime regulation.

The European Union has been at the forefront of regulatory innovation. The AI Act, which entered into force on August 1, 2024, classifies AI systems by risk level and sets requirements for transparency and labeling of deepfake content. It defines deepfakes as "AI-generated or AI-altered content (image, audio, or video) that may mislead a person into believing it is authentic" and requires both human-recognizable markings and machine-readable watermarking to authenticate provenance [28].

Another milestone is the Digital Services Act (DSA), in force since 2022, which obliges online platforms – particularly Very Large Online Platforms (VLOPs) – to improve algorithmic transparency, promptly detect and remove illegal content, and address deepfake risks. The European Commission has already requested companies such as Google, Facebook, and TikTok to submit plans for mitigating generative AI risks, with the DSA providing enforcement tools, including fines [29].

National initiatives are also emerging. Spain has introduced a bill imposing significant fines – up to €35 million or 7% of global revenue – for failure to label AI-generated content, in alignment with the AI Act [30]. Denmark has proposed amendments to its copyright law, granting citizens explicit rights over their image and voice, thereby allowing fines against platforms that fail to remove deepfake content upon notification [31].

The United Kingdom has pursued a more comprehensive approach to digital governance. The Online Safety Act, adopted in 2023, is one of the most ambitious pieces of legislation in the field of digital security. It places extensive obligations on platforms and service providers to identify, restrict, and remove harmful content, including AI-generated materials [32].

A key feature of the Act is the imposition of obligations on digital services to prevent the dissemination of harmful and fraudulent content. Unlike some jurisdictions, the UK framework emphasizes not only corporate liability but also the personal liability of company executives for non-compliance with regulatory requirements. This model considerably strengthens accountability and encourages companies to adopt monitoring and filtering mechanisms aimed at limiting the circulation of deepfake content.

In addition, the Act establishes demanding transparency rules for both content moderation processes and recommendation algorithms. This is especially important because the rapid spread of synthetic audio-visual materials often depends on algorithm-driven amplification. In this context, openness and accountability are regarded as vital tools for reinforcing trust in the digital space and for reducing the risk of large-scale distribution of deepfakes.

Lastly, the Online Safety Act adopts a preventive stance: its focus lies in lowering the probability of misuse before it occurs, rather than simply reacting once violations have taken place. In this regard, the UK framework illustrates a forward-looking strategy that gives priority to user protection and the building of a more resilient digital ecosystem.

The U.S. legal system approaches deepfake-related abuses primarily through the adaptation of existing criminal law mechanisms, without yet establishing a separate offense specific to synthetic media. The main tools applied are statutes on wire fraud and identity theft, which have traditionally addressed deception carried out via electronic communications [33].

The category of wire fraud encompasses a broad range of unlawful activities involving electronic means for deception and unlawful enrichment. In the context of deepfakes, this provision applies to fraudulent video calls impersonating executives, the distribution of fabricated audio and video materials, and investment schemes based on synthetic representations of public figures. The flexibility of the statute's wording enables prosecutors to incorporate emerging forms of deception made possible by generative AI technologies.

Identity theft legislation plays an equally significant role. It criminalizes the unlawful use of identifying data – such as names, likenesses, or voices – which directly aligns with deepfake practices. U.S. enforcement prioritizes protecting individuals from the misuse of personal data for fraudulent or defamatory purposes, allowing prosecutors to classify deepfake-based conduct as identity theft even without explicit statutory references to synthetic media.

Further developments are reflected in regulatory activity and state-level initiatives. In California, specific rules now limit the use of deepfakes in both election campaigns and the production of pornographic materials without consent. Measures of this kind reflect a broader movement toward targeted regulation aimed at protecting society as a whole while also safeguarding individual rights.

The U.S. approach can be seen as a pragmatic compromise: it relies on adaptable general provisions while gradually adding more specialized norms. This enables regulators to react to emerging risks without unduly constraining technological progress. At the same time, the lack of a federal law dealing directly with deepfakes leads to uneven enforcement and potential inconsistencies between states.

China, on the other hand, was among the first countries to put in place a dedicated regulatory regime for synthetic media. In 2022, the Cyberspace Administration of China (CAC) adopted the Administrative Provisions on Deep Synthesis, which entered into force on January 10, 2023 [34].

A central element of this framework is the mandatory labeling of synthetic content. Creators and distributors of deepfake materials must indicate that the product was generated using deep synthesis technologies, thereby reducing the risk of misleading users [35]. The provisions also oblige services to remove content infringing individual rights or undermining public order.

The regulations assign significant responsibility to providers through the principle of strict or vicarious liability, requiring platforms and services to implement internal monitoring, user verification, content and algorithm management, and removal of unlawful materials.

Moreover, China's legal model reflects a preventive orientation, emphasizing public stability and information security rather than a balance between innovation and user rights, which distinguishes it from Western approaches. While this facilitates swift responses to fraudulent or defamatory content, it also raises concerns about excessive state control over digital innovation.

In 2023, China further expanded its regulatory framework by issuing the Interim Measures for the Management of Generative AI Services, extending obligations and liability to providers of generative AI services [36].

Kazakhstan's legislation currently lacks specific provisions directly regulating the use of deepfake technologies. Nevertheless, certain norms of criminal and information law partially address this domain. The primary tool for criminal qualification is Article 190 of the Criminal Code of the Republic of Kazakhstan, which establishes liability for misappropriation of property or acquisition of property rights through deception or abuse of trust (Part 1), and for fraud "through deception or abuse of trust of an information system user" (Part 2) [37]. This formulation reflects a gradual adaptation of criminal law to the digital environment and may be applied in cases involving fabricated audio or video materials. Yet the absence of explicit references to synthetic media creates uncertainty in law enforcement practice and complicates the qualification of crimes involving generative technologies.

At present, there is no separate criminal liability for creating deepfake content in Kazakhstan. However, certain provisions of the Criminal and Administrative Codes already cover its consequences – fraud, reputational harm, and the use of information technologies as aggravating circumstances. In this context, members of the Mazhilis have proposed two possible legislative pathways: first, to amend Article 190 of the Criminal Code by recognizing the use of AI and deepfakes as aggravating factors; second, to introduce a distinct offense for the creation and dissemination of synthetic content [38].

The jurisdictional problems create additional obstacles for authorities to address. The current territorial and personal jurisdiction rules fail to address deepfake cybercrimes because servers and victims and perpetrators exist in separate national territories. The situation requires international treaties and conventions to establish universal mechanisms for handling these cases. The Budapest Convention on Cybercrime from 2001 contains no specific rules about synthetic media, which makes its practical value limited.

The Administrative Code of Russia, through Article 456-2 and the Criminal Code of Russia through Article 274, establishes two separate legal frameworks to handle the distribution of intentionally false information, which allows content blocking and data transfer to law enforcement. Nonetheless, these mechanisms do not treat deepfakes as a distinct category of crime [39]. Enhanced penalties under Article 190, depending on circumstances (organized character, abuse of official position, large-scale or especially large-scale damages), illustrate an effort to account for the degree of public harm. Yet the potential of deep synthesis technologies to inflict substantial damage with minimal resources necessitates the creation of specialized norms.

The protection of personal rights against unlawful use of data is also governed by the Law of the Republic of Kazakhstan "On Personal Data and Its Protection" [40], which defines personal data as any information identifying an individual, including images and voice. Accordingly, the creation and dissemination of deepfake content without consent violates the principle of lawful data processing. However, the current law does not clarify the status of falsified biometric data, leaving citizens insufficiently protected from abuses linked to synthetic media.

Judicial practice also faces serious challenges. In the absence of procedural standards for authenticating digital evidence, courts rely on expert evaluations, whose methodologies lack sufficient unification, thereby weakening prosecution effectiveness and generating inconsistent outcomes [41]. The digital transformation and increasing synthetic media adoption in Kazakhstan require immediate action to update national laws that address emerging challenges. The protection of individual rights and judicial consistency in deepfake-related crimes will improve when specific laws establish responsibility for fake content creation and synthetic biometric data receives a clear legal classification.

A complete solution to deepfake misuse requires organizations to unite technological progress with legal frameworks and educational programs. The fight against new online scams requires joint efforts between government bodies and academic institutions and private businesses and civil organizations.

On the technological side, the main focus should be on developing effective tools for identifying and labeling synthetic materials. Techniques such as digital watermarking and provenance verification make it possible to register the creation of content and to trace subsequent alterations. While modern detection methods based on micro-distortions in images, voice, and motion dynamics are improving, they require continuous adaptation to rapidly advancing generative models. The implementation of such technologies across digital platforms and communication services could significantly strengthen resilience against fraudulent schemes.

From a regulatory standpoint, criminal and procedural law requires refinement. Introducing provisions directly establishing liability for the creation and dissemination of synthetic media for fraudulent or harmful purposes would reduce legal uncertainty and enhance the qualification of offenses. Equally important is the establishment of procedural rules for authenticating digital evidence, necessary for ensuring admissibility in judicial proceedings. In addition, harmonization of national approaches with international standards is critical to enable information exchange and joint investigation of cross-border crimes.

The educational dimension is no less vital, as the success of deepfake-enabled attacks often depends on victims' awareness and psychological resilience. Improving digital literacy, teaching recognition of manipulative social engineering strategies, and fostering critical information evaluation can reduce fraud effectiveness. In corporate settings, such measures should include training sessions, simulated phishing and deepfake attacks, and the introduction of internal verification protocols.

The integration of technological, regulatory, and educational measures thus provides the foundation for a systemic response to crimes involving deep synthesis technologies. Only through their comprehensive application can society achieve a balance between technological innovation and protection against abuses in the digital environment.

Conclusion

With the fast-paced advancement of artificial intelligence – and generative neural networks in particular – deepfake technologies have moved beyond their original use in the entertainment sphere and become a serious tool for criminal activity, now regarded as one of the major risks in the digital landscape. The findings of this study demonstrate that the combination of deepfakes with social engineering techniques creates multidimensional threats capable of destabilizing financial systems, compromising information security, damaging reputations, and infringing upon individual rights. The manifestations of such threats are diverse and include large-scale corporate fraud based on executive impersonation, voice cloning attacks targeting biometric authentication systems, fraudulent investment schemes promoted through fabricated endorsements, and the dissemination of intimate synthetic content used for blackmail. The high realism and accessibility of these technologies lower the threshold for offenders and enhance the effectiveness of psychological manipulation, making individuals vulnerable even when they are generally aware of potential risks.

The challenges posed by deepfake-enabled crimes expose fundamental limitations of existing legal systems, which were originally designed to address traditional forms of fraud and identity misuse. Among the most pressing concerns are the ambiguous legal qualification of synthetic media–related offenses, the complexities of authenticating digital evidence, the transnational character of such crimes, and the unresolved issue of liability for platform operators. International experience demonstrates a variety of approaches: the European Union has adopted the AI Act, introducing requirements for transparency and labeling of synthetic content; China has implemented Deep Synthesis Provisions with similar objectives; the United States relies on adapting existing legal instruments; while the United Kingdom has enacted the Online Safety Act, combining regulatory oversight with user protection mechanisms. These examples illustrate the necessity of a comprehensive and multi-level approach.

For Kazakhstan, the research emphasizes the need to modernize national legislation to explicitly regulate the misuse of synthetic media. Possible directions include amending Article 190 of the Criminal Code to recognize the use of deepfakes as an aggravating factor in fraud cases, or the introduction of a distinct criminal provision addressing harmful applications of synthetic content. Equally important is the development of legal norms protecting falsified biometric data, such as voice or facial images, as well as the establishment of procedural standards for authenticating and admitting synthetic media as evidence in judicial proceedings.

Institutional practices must also be adapted to meet the new challenges. Digital platforms should be obliged to ensure the labeling and timely removal of manipulated content, with a clear delineation of liability between operators and direct perpetrators. Financial and other organizations, in turn, need to enhance resilience against deepfake-enabled fraud by adopting multi-level approval mechanisms for sensitive operations and introducing technological safeguards that combine biometric and behavioral verification methods.

Finally, given the inherently cross-border nature of these threats, effective counteraction requires the harmonization of national legal frameworks with international standards and the creation of mechanisms for the recognition and exchange of digital evidence across jurisdictions.

In conclusion, the study demonstrates that the risks posed by deepfake technologies and social engineering cannot be addressed in isolation. A comprehensive response that integrates legal reforms, institutional resilience, technological innovation, and international cooperation is necessary. Only such an integrated and systemic approach will enable states to maintain a sustainable balance between fostering technological innovation and protecting society from the harmful exploitation of generative technologies.

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26103625 «Online fraud using deepfake-

technologies and social engineering: problems of criminal law counteraction, prospects for legislative regulation»).

The contribution of the authors

A.B. Smanova – corresponding author, developed the overall concept of the research, formulated the scientific problem, and coordinated the preparation of the article.

A.Zh. Muratova – carried out the collection and analysis of materials, prepared the section on mechanisms of social engineering and their role in online fraud, and participated in drafting the text.

Sh.R. Zhumagulova – conducted the review of legislative and regulatory frameworks, analyzed legal challenges of combating deepfake technologies in fraud, and finalized the conclusions.

References

1 Хасанай А., Абылайулы А. Ответственность в контексте применения искусственного интеллекта в условиях вооруженных конфликтов // Вестник Евразийского национального университета имени Л.Н. Гумилева. Серия: Право. 2024. Т.148. №3. С. 55–74. https://doi.org/10.32523/2616-6844-2024-148-3-55-74

2 Mashaev A. The dark side of AI: Assessing the top cyber threats to Kazakhstan. – Режим доступа: https://kz.kursiv.media/en/2025-08-12/engk-yeri-the-dark-side-of-ai-assessing-the-top-cyber-threats-to-kazakhstan/ (дата обращения: 20.05.2025).

3 Deepfake Fraud Costs the Financial Sector an Average of $600,000 for Each Company. – Режим доступа: https://regulaforensics.com/news/deepfake-fraud-costs/ (дата обращения: 15.11.2024).

4 Colman B. Why detecting dangerous AI is key to keeping trust alive in the deepfake era. – Режим доступа:https://www.weforum.org/stories/2025/07/why-detecting-dangerous-ai-is-key-to-keeping-trust-alive/ (дата обращения: 15.04.2025).

5 Pedersen K.T., Pepke L., Stærmose T., Papaioannou M., Choudhary G., Dragoni N. Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments // J. Cybersecur. Priv. 2025. Vol.5. No2. P. 18. https://doi.org/10.3390/jcp5020018

6 Dunsin D. Deepfake and Biometric Spoofing: AI-Driven Identity Fraud and Countermeasures. 2025. – Режим доступа: https://www.researchgate.net/publication/390141504_Deepfake_and_Biometric_Spoofing_AI-Driven_Identity_Fraud_and_Countermeasures (дата обращения: 20.03.2025).

7 Lalchand S., Srinivas V., Maggiore B., Henderson J. Generative AI is expected to magnify the risk of deepfakes and other fraud in banking. – Режим доступа: https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html (дата обращения: 01.06.2024).

8 Deepfakes in the Financial Sector: Understanding the Threats, Managing the Risks. A Report by the FS-ISAC Artificial Intelligence Risk Working Group. – Режим доступа: https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf (дата обращения: 10.11.2024).

9 Robins-Early N. CEO of world’s biggest ad firm targeted by deepfake scam. –Режим доступа: https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam (дата обращения: 20.05.2024).

10 Forrest D. Challenges in voice biometrics: Vulnerabilities in the age of deepfakes. – Режим доступа: https://bankingjournal.aba.com/2024/02/challenges-in-voice-biometrics-vulnerabilities-in-the-age-of-deepfakes/ (дата обращения: 25.02.2024).

11 Рождайкина Е.И. Проблемы защиты биометрических персональных данных при расследовании преступлений // Вопросы российской юстиции. 2023. №28. С. 378–386.

12 Popa C., Kesavarajah A., Tahiri H., Cunningham L., Pallath R., Wu T. Deepfake Technology Unveiled: The Commoditization of AI and Its Impact on Digital Trust. – Режим доступа: https://www.arxiv.org/pdf/2506.07363 (дата обращения: 25.05.2025).

13 Alexander A. The New Identity Theft: Deepfakes and the Rise of Synthetic Impersonation Scams. 2025. https://doi.org/10.2139/ssrn.5368947

14 Romero-Moreno F. Deepfake detection in generative AI: A legal framework proposal to protect human rights // Computer Law & Security Review. 2025. Vol.58. Pp. 106–162. ISSN 2212-473X. https://doi.org/10.1016/j.clsr.2025.106162

15 Han C., Li A., Kumar D., Durumeric Z. Characterizing the MrDeepFakes Sexual Deepfake Marketplace. – Режим доступа: https://arxiv.org/abs/2410.11100 (дата обращения: 15.03.2025).

16 Blancaflor E., Garcia J.I., Magno F.D., Vilar M.J. Deepfake Blackmailing on the Rise: The Burgeoning Posterity of Revenge Pornography in the Philippines // ICIIT '24: Proceedings of the 9th International Conference on Intelligent Information Technology. 2024. Pp. 295–301. https://doi.org/10.1145/3654522.365454

17 Arya C.S. et al. A Review Paper on Developing a Real-Time Deepfake Voice Synthesis Framework: A Study in Artificial Intelligence // Educational Administration: Theory and Practice. 2024. Vol.30. No4. Pp. 1455–1461. https://doi.org/10.53555/kuey.v30i4.1692

18 Blake H. AI-Powered Social Engineering: Understanding the Role of Deepfake Technology in Exploiting Human Trust. – Режим доступа: https://www.researchgate.net/publication/388931016_AI-Powered_Social_Engineering_Understanding_the_Role_of_Deepfake_Technology_in_Exploiting_Human_Trust (дата обращения: 10.04.2025).

19 Dunsin D. Deepfake Technology and AI-Driven Social Engineering Attacks: Implications for Cyber Defense. – Режим доступа: https://www.researchgate.net/publication/390665794_Deepfake_Technology_and_AI-Driven_Social_Engineering_Attacks_Implications_for_Cyber_Defense(дата обращения: 20.06.2022).

20 Yu J., Yu Y., Wang X., Lin Y., Yang M., Qiao Y., Wang F.-Y. The Shadow of Fraud: The Emerging Danger of AI-powered Social Engineering and its Possible Cure. – Режим доступа: https://arxiv.org/abs/2407.15912 (дата обращения: 15.09.2024).

21 Selvamuthukumaran V. Advancing Deepfake Legislation: Comparative Analysis and Pathways for Policy Change. – Режим доступа: https://criticaldebateshsgj.scholasticahq.com/post/3143-advancing-deepfake-legislation-comparative-analysis-and-pathways-for-policy-change-by-virthiha-selvamuthukumaran (дата обращения: 25.05.2025).

22 Bhattathiri A., Sharma F., Purohit A. Deepfake in the Courtroom: Legal Challenges and Evidentiary Standards // Digital Doppelgangers: The Rise of Deepfakes & Artificial Intelligence. Lex Assisto Media and Publications. 2025. Pp. 10–24. – Режим доступа: https://www.researchgate.net/publication/390200521_Deepfake_in_the_Courtroom_Legal_Challenges_and_Evidentiary_Standards (дата обращения: 20.02.2025).

23 Amerini I. et al. Deepfake Media Forensics: State of the Art and Challenges Ahead. – Режим доступа: https://arxiv.org/abs/2408.00388 (дата обращения: 10.10.2024).

24 LaMonaga J.P. A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes // American University Law Review. 2020. Vol.69. Iss.6. Article 5. Pp. 1942–1988.

25 Ma Y. Deepfake Policy Brief. 2021. https://doi.org/10.2139/ssrn.5038837

26 Khan F. Does the Digital Services Act achieve a balance between regulating deepfakes and protecting the fundamental right to freedom of expression? 2024. http://dx.doi.org/10.2139/ssrn.4868290

27 Fragale M., Grilli V. Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation. – Режим доступа: https://cjel.law.columbia.edu/preliminary-reference/2024/deepfake-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/ (дата обращения: 15.12.2024).

28 Fritz G., Ehlen T., Fokter Cuvan T. EU AI Act unpacked #8: New rules on deepfakes. – Режим доступа:https://technologyquotient.freshfields.com/post/102jb19/eu-ai-act-unpacked-8-new-rules-on-deepfakes (дата обращения: 15.07.2024).

29 Chan K. Europe asks Google, Facebook, TikTok and other platforms how they're reducing generative AI risks. – Режим доступа: https://apnews.com/article/generative-ai-risks-digital-services-act-europe-1cc677bbbfaa919a5f309fbd5ccfedb7 (дата обращения: 25.03.2024).

30 Spain to impose massive fines for not labelling AI-generated content. – Режим доступа: https://www.reuters.com/technology/artificial-intelligence/spain-impose-massive-fines-not-labelling-ai-generated-content-2025-03-11/ (дата обращения: 20.03.2025).

31 Jaiswal N. 'Your face, your rights': Denmark's tough new Deepfake law could change how AI imitations are handled across Europe. – Режим доступа: https://www.indiatimes.com/news/your-face-your-rights-denmarks-tough-new-deepfake-law-could-change-how-ai-imitations-are-handled-across-europe/articleshow/122759842.html (дата обращения: 15.06.2025).

32 Ofcom's approach to implementing the Online Safety Act. – Режим доступа: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation (дата обращения: 15.11.2023).

33 Eichner A.W. Artificial Intelligence and Weaponized Illusions: Methodologies for Federal Fraud Prosecutions Involving Deepfakes // American University Law Review. 2024. Vol.73. Iss.5. Article 2. Pp. 1317–1366.

34 Filipova I.A. Legal Regulation of Artificial Intelligence: Experience of China // Journal of Digital Technologies and Law. 2024. Vol.2. No1. Pp. 46–73. https://doi.org/10.21202/jdtl.2024.4

35 Zou M., Zhang L. Navigating China's regulatory approach to generative artificial intelligence and large language models // Cambridge Forum on AI: Law and Governance. 2025. Vol.1. e8. https://doi.org/10.1017/cfl.2024.4

36 China: Data and evolving digital regulation: algorithm regulation. – Режим доступа: https://www.twobirds.com/en/capabilities/practices/digital-rights-and-assets/apac-dra/apac-dsd/data-as-a-key-digital-asset/china/data-and-evolving-digital-regulation-algorithm-regulation (дата обращения: 15.01.2024).

37 Penal Code of the Republic of Kazakhstan. The Code of the Republic of Kazakhstan dated 3 July 2014 №226-V of the Law of the Republic of Kazakhstan. – Режим доступа: https://adilet.zan.kz/eng/docs/K1400000226 (дата обращения: 15.02.2025).

38 Уголовную ответственность за дипфейки могут ввести в Казахстане. – Режим доступа: https://bluescreen.kz/ugholovnuiu-otvietstviennost-za-dipfieiki-moghut-vviesti-v-kazakhstanie/ (дата обращения: 20.05.2025).

39 Кусаинова И. Число уголовных дел за дипфейки растет в Казахстане. – Режим доступа: https://www.inbusiness.kz/ru/news/chislo-ugolovnyh-del-za-dipfejki-rastet-v-kazahstane (дата обращения: 25.05.2025).

40 The Law of the Republic of Kazakhstan dated 21 May 2013 №94-V On Personal Data and their Protection. – Режим доступа: https://adilet.zan.kz/eng/docs/Z1300000094 (дата обращения: 15.02.2025).

41 Мицкая Е.В. Вопросы правового противодействия технологии deepfake // Российско-азиатский правовой журнал. 2025. №1. С. 50–58. https://doi.org/10.14258/ralj(2025)1.8

А.Б. Сманова[1], А.Ж. Муратова[2], Ш.Р. Жумагулова[3]
[1]Әл-Фараби Қазақ Ұлттық университеті, Алматы, Қазақстан
[2]Мұхамеджан Тынышбаев атындағы ALT университеті, Алматы, Қазақстан
[3]Қорқыт Ата атындағы Қызылорда университеті, Қызылорда, Қазақстан
(e-mail: [1]akmaralbahtyar@gmail.com, [2]kkaebsong98@mail.ru, [3]jumagulova_sholpan@mail.ru)

Онлайн алаяқтықтағы дипфейк технологиялары мен әлеуметтік инженерия: нысандары, тетіктері және құқықтық мәселелері

Аңдатпа: Жасанды интеллекттің, әсіресе генеративті нейрондық желілердің жедел дамуы цифрлық ортада сапалық жаңа қатерлердің пайда болуына әкелді. Дипфейк технологиялары шынайы материалдардан ажырату қиын синтетикалық аудио және бейнемазмұнды жасауға мүмкіндік береді. Олар әлеуметтік инженерия әдістерімен біріктірілгенде, алаяқтық схемалардың нанымдылығын күшейтіп, елеулі қаржылық шығындарға және киберқылмыстың тұрақты өсуіне алып келеді. Қазіргі қолданыстағы құқықтық тетіктер көптеген елдерде, соның ішінде Қазақстанда, бұл сын-қатерлерге тиімді қарсы тұру үшін жеткіліксіз болып отыр.

Зерттеудің мақсаты – дипфейктер мен әлеуметтік инженерияның онлайн-алаяқтықтағы қолданыс нысандары мен механизмдерін кешенді талдау, қарсы іс-қимылға кедергі келтіретін құқықтық мәселелерді анықтау және заңнаманы жетілдіру бойынша ұсынымдар әзірлеу.

Жұмыстың ғылыми және практикалық маңызы – психологиялық манипуляция мен генеративті жасанды интеллект арасындағы өзара байланысты ашуда, сондай-ақ киберқауіптерге қарсы ұлттық стратегияны қалыптастыруға бағытталған ұсыныстар ұсынуда. Әдіснамалық негізі ретінде синтетикалық медианың қылмыстық қолданылуының нақты мысалдарын зерттеу, халықаралық және ұлттық заңнаманы салыстырмалы талдау, алдын алу шараларын бағалау кіретін кешенді аналитикалық тәсіл қолданылды.

Зерттеу нәтижелері дипфейктердің жетекшілерді имитациялау арқылы қаржылық алаяқтықта, дауыс биометриясын айналып өтуде, қоғамдық тұлғалардың жалған бейнелерін пайдаланып инвестициялық алаяқтықта және құпиялылықты бұзу мен бопсалауда кеңінен қолданылатынын көрсетті. Негізгі құқықтық проблемаларға қылмысты саралау қиындығы, цифрлық дәлелдемелердің түпнұсқалығын айқындау, кибершабуылдардың трансшекаралық сипаты және қылмыскерлер мен онлайн-платформалар арасындағы жауапкершілікті бөлу жатады.

Зерттеу қорытындысы интеграцияланған тәсілдің қажеттілігін айқындайды: синтетикалық медианы анықтау және таңбалау құралдарын әзірлеу, қылмыстық және іс жүргізу заңнамасын жетілдіру, халықаралық стандарттарды үйлестіру, сондай-ақ халықтың цифрлық сауаттылығы мен психологиялық тұрақтылығын арттыру. Бұл жұмыс терең синтез технологияларын пайдаланатын киберқылмысқа жүйелі жауап қалыптастыруға арналған теориялық және практикалық негізді ұсынады.

Түйін сөздер: дипфейк-технологиялар, әлеуметтік инженерия, онлайн алаяқтық, жасанды интеллект, құқықтық мәселелер, киберқылмыс.

А.Б. Сманова[1], А.Ж. Муратова[2], Ш.Р. Жумагулова[3]

[1]Казахский Национальный университет имени аль-Фараби, Алматы, Казахстан
[2]ALT Университет имени Мухамеджана Тынышпаева, Алматы, Казахстан
[3]Кызылординский университет имени Коркыт Ата, Кызылорда, Казахстан
(e-mail: [1]akmaralbahtyar@gmail.com, [2]kkaebsong98@mail.ru, [3]jumagulova_sholpan@mail.ru)

Технологии дипфейка и социальная инженерия в онлайн-мошенничестве: формы, механизмы и правовые вызовы

Аннотация: Бурное развитие технологий искусственного интеллекта, в частности генеративных нейронных сетей, обусловило появление качественно новых угроз в цифровой среде. Особенно актуальной стала проблема распространения дипфейк-технологий, позволяющих создавать синтетические аудио- и видеоматериалы, практически неотличимые от подлинных. Их использование в сочетании с методами социальной инженерии усиливает убедительность мошеннических схем, что приводит к существенным финансовым потерям и росту киберпреступности. Существующие правовые механизмы в большинстве стран, включая Казахстан, остаются недостаточными для эффективного противодействия указанным вызовам.

Целью исследования является комплексный анализ форм и механизмов применения дипфейков и социальной инженерии в онлайн-мошенничестве, выявление ключевых правовых проблем, препятствующих борьбе с ними, а также разработка предложений по совершенствованию законодательства и институциональных практик.

Научная новизна и практическая значимость работы заключаются в выявлении взаимосвязи между психологическими манипуляциями и генеративным ИИ, а также в формулировании рекомендаций для формирования национальной стратегии противодействия киберугрозам. Методологическая база исследования основана на комплексном аналитическом подходе, включающем изучение примеров преступного применения синтетических медиа, сравнительный анализ международного и национального законодательства, а также оценку превентивных мер.

Результаты исследования показали, что дипфейки используются для финансовых махинаций с имитацией руководителей компаний, обхода систем биометрической аутентификации, инвестиционного мошенничества с использованием образов публичных лиц, а также для нарушения конфиденциальности и вымогательства. Основными правовыми проблемами являются сложности квалификации таких преступлений, установления подлинности цифровых доказательств, трансграничный характер кибератак и неопределённость распределения ответственности между злоумышленниками и цифровыми платформами.

Заключение подтверждает необходимость интеграции технологических, правовых и образовательных мер: разработка инструментов обнаружения и маркировки синтетических медиа, совершенствование уголовного и процессуального законодательства, гармонизация международных норм, а также повышение уровня цифровой грамотности населения. Вклад исследования заключается в формировании теоретических и практических основ для создания системного ответа на угрозы, связанные с технологиями глубокого синтеза.

Ключевые слова: дипфейк-технологии, социальная инженерия, онлайн-мошенничество, искусственный интеллект, правовые вызовы, киберпреступность.

References

1 Khasanai A., Abylaiuly A. Otvetstvennost' v kontekste primeneniya iskusstvennogo intellekta v usloviyah vooruzhennyh konfliktov // Vestnik Evrazijskogo nacional'nogo universiteta imeni L.N. Gumileva. Seriya: Pravo. – 2024. – T.148. – No3. – S. 55–74. – https://doi.org/10.32523/2616-6844-2024-148-3-55-74

2 Mashaev A. The dark side of AI: Assessing the top cyber threats to Kazakhstan. – URL: https://kz.kursiv.media/en/2025-08-12/engk-yeri-the-dark-side-of-ai-assessing-the-top-cyber-threats-to-kazakhstan/ (data obrashcheniya: 20.05.2025).

3 Deepfake Fraud Costs the Financial Sector an Average of $600,000 for Each Company. – URL: https://regulaforensics.com/news/deepfake-fraud-costs/ (data obrashcheniya: 15.11.2024).

4 Colman B. Why detecting dangerous AI is key to keeping trust alive in the deepfake era. – URL: https://www.weforum.org/stories/2025/07/why-detecting-dangerous-ai-is-key-to-keeping-trust-alive/ (data obrashcheniya: 15.04.2025).

5 Pedersen K.T., Pepke L., Stærmose T., Papaioannou M., Choudhary G., Dragoni N. Deepfake-Driven Social Engineering: Threats, Detection Techniques, and Defensive Strategies in Corporate Environments // J. Cybersecur. Priv. – 2025. – Vol.5. – No2. – P. 18. – https://doi.org/10.3390/jcp5020018

6 Dunsin D. Deepfake and Biometric Spoofing: AI-Driven Identity Fraud and Countermeasures. – 2025. – URL: https://www.researchgate.net/publication/390141504_Deepfake_and_Biometric_Spoofing_AI-Driven_Identity_Fraud_and_Countermeasures (data obrashcheniya: 20.03.2025).

7 Lalchand S., Srinivas V., Maggiore B., Henderson J. Generative AI is expected to magnify the risk of deepfakes and other fraud in banking. – URL: https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html (data obrashcheniya: 01.06.2024).

8 Deepfakes in the Financial Sector: Understanding the Threats, Managing the Risks. A Report by the FS-ISAC Artificial Intelligence Risk Working Group. – URL: https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf (data obrashcheniya: 10.11.2024).

9 Robins-Early N. CEO of world's biggest ad firm targeted by deepfake scam. – URL: https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam (data obrashcheniya: 20.05.2024).

10 Forrest D. Challenges in voice biometrics: Vulnerabilities in the age of deepfakes. – URL: https://bankingjournal.aba.com/2024/02/challenges-in-voice-biometrics-vulnerabilities-in-the-age-of-deepfakes/ (data obrashcheniya: 25.02.2024).

11 Rozhdajkina E.I. Problemy zashchity biometricheskih personal'nyh dannyh pri rassledovanii prestuplenij // Voprosy rossijskoj yusticii. – 2023. – No28. – S. 378–386.

12 Popa C., Kesavarajah A., Tahiri H., Cunningham L., Pallath R., Wu T. Deepfake Technology Unveiled: The Commoditization of AI and Its Impact on Digital Trust. – URL: https://www.arxiv.org/pdf/2506.07363 (data obrashcheniya: 25.05.2025).

13 Alexander A. The New Identity Theft: Deepfakes and the Rise of Synthetic Impersonation Scams. – 2025. – https://doi.org/10.2139/ssrn.5368947

14 Romero-Moreno F. Deepfake detection in generative AI: A legal framework proposal to protect human rights // Computer Law & Security Review. – 2025. – Vol.58. – Pp. 106–162. – ISSN 2212-473X. – https://doi.org/10.1016/j.clsr.2025.106162

15 Han C., Li A., Kumar D., Durumeric Z. Characterizing the MrDeepFakes Sexual Deepfake Marketplace. – URL: https://arxiv.org/abs/2410.11100 (data obrashcheniya: 15.03.2025).

16 Blancaflor E., Garcia J.I., Magno F.D., Vilar M.J. Deepfake Blackmailing on the Rise: The Burgeoning Posterity of Revenge Pornography in the Philippines // ICIIT '24: Proceedings of the 9th International Conference on Intelligent Information Technology. – 2024. – Pp. 295–301. – https://doi.org/10.1145/3654522.365454

17 Arya C.S. et al. A Review Paper on Developing a Real-Time Deepfake Voice Synthesis Framework: A Study in Artificial Intelligence // Educational Administration: Theory and Practice. – 2024. – Vol.30. – No4. – Pp. 1455–1461. – https://doi.org/10.53555/kuey.v30i4.1692

18 Blake H. AI-Powered Social Engineering: Understanding the Role of Deepfake Technology in Exploiting Human Trust. – URL: https://www.researchgate.net/publication/388931016_AI-Powered_Social_Engineering_Understanding_the_Role_of_Deepfake_Technology_in_Exploiting_Human_Trust (data obrashcheniya: 10.04.2025).

19 Dunsin D. Deepfake Technology and AI-Driven Social Engineering Attacks: Implications for Cyber Defense. – URL: https://www.researchgate.net/publication/390665794_Deepfake_Technology_and_AI-Driven_Social_Engineering_Attacks_Implications_for_Cyber_Defense (data obrashcheniya: 20.06.2022).

20 Yu J., Yu Y., Wang X., Lin Y., Yang M., Qiao Y., Wang F.-Y. The Shadow of Fraud: The Emerging Danger of AI-powered Social Engineering and its Possible Cure. – URL: https://arxiv.org/abs/2407.15912 (data obrashcheniya: 15.09.2024).

21 Selvamuthukumaran V. Advancing Deepfake Legislation: Comparative Analysis and Pathways for Policy Change. – URL: https://criticaldebateshsgj.scholasticahq.com/post/3143-advancing-deepfake-legislation-comparative-analysis-and-pathways-for-policy-change-by-virthiha-selvamuthukumaran (data obrashcheniya: 25.05.2025).

22 Bhattathiri A., Sharma F., Purohit A. Deepfake in the Courtroom: Legal Challenges and Evidentiary Standards // Digital Doppelgangers: The Rise of Deepfakes & Artificial Intelligence. – Lex Assisto Media and Publications. – 2025. – Pp. 10–24. – URL: https://www.researchgate.net/publication/390200521_Deepfake_in_the_Courtroom_Legal_Challenges_and_Evidentiary_Standards (data obrashcheniya: 20.02.2025).

23 Amerini I. et al. Deepfake Media Forensics: State of the Art and Challenges Ahead. – URL: https://arxiv.org/abs/2408.00388 (data obrashcheniya: 10.10.2024).

24 LaMonaga J.P. A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes // American University Law Review. – 2020. – Vol.69. – Iss.6. – Article 5. – Pp. 1942–1988.

25 Ma Y. Deepfake Policy Brief. – 2021. – https://doi.org/10.2139/ssrn.5038837

26 Khan F. Does the Digital Services Act achieve a balance between regulating deepfakes and protecting the fundamental right to freedom of expression? – 2024. – http://dx.doi.org/10.2139/ssrn.4868290

27 Fragale M., Grilli V. Deepfake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation. – URL: https://cjel.law.columbia.edu/preliminary-reference/2024/deepfake-deep-trouble-the-european-ai-act-and-the-fight-against-ai-generated-misinformation/ (data obrashcheniya: 15.12.2024).

28 Fritz G., Ehlen T., Fokter Cuvan T. EU AI Act unpacked #8: New rules on deepfakes. – URL: https://technologyquotient.freshfields.com/post/102jb19/eu-ai-act-unpacked-8-new-rules-on-deepfakes (data obrashcheniya: 15.07.2024).

29 Chan K. Europe asks Google, Facebook, TikTok and other platforms how they're reducing generative AI risks. – URL: https://apnews.com/article/generative-ai-risks-digital-services-act-europe-1cc677bbbfaa919a5f309fbd5ccfedb7 (data obrashcheniya: 25.03.2024).

30 Spain to impose massive fines for not labelling AI-generated content. – URL: https://www.reuters.com/technology/artificial-intelligence/spain-impose-massive-fines-not-labelling-ai-generated-content-2025-03-11/ (data obrashcheniya: 20.03.2025).

31 Jaiswal N. 'Your face, your rights': Denmark's tough new Deepfake law could change how AI imitations are handled across Europe. – URL: https://www.indiatimes.com/news/your-face-your-rights-denmarks-tough-new-deepfake-law-could-change-how-ai-imitations-are-handled-across-europe/articleshow/122759842.html (data obrashcheniya: 15.06.2025).

32 Ofcom's approach to implementing the Online Safety Act. – URL: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/roadmap-to-regulation (data obrashcheniya: 15.11.2023).

33 Eichner A.W. Artificial Intelligence and Weaponized Illusions: Methodologies for Federal Fraud Prosecutions Involving Deepfakes // American University Law Review. – 2024. – Vol.73. – Iss.5. – Article 2. – Pp. 1317–1366.

34 Filipova I.A. Legal Regulation of Artificial Intelligence: Experience of China // Journal of Digital Technologies and Law. – 2024. – Vol.2. – No1. – Pp. 46–73. – https://doi.org/10.21202/jdtl.2024.4

35 Zou M., Zhang L. Navigating China's regulatory approach to generative artificial intelligence and large language models // Cambridge Forum on AI: Law and Governance. – 2025. – Vol.1. – e8. – https://doi.org/10.1017/cfl.2024.4

36 China: Data and evolving digital regulation: algorithm regulation. – URL: https://www.twobirds.com/en/capabilities/practices/digital-rights-and-assets/apac-dra/apac-dsd/data-as-a-key-digital-asset/china/data-and-evolving-digital-regulation-algorithm-regulation (data obrashcheniya: 15.01.2024).

37 Penal Code of the Republic of Kazakhstan. The Code of the Republic of Kazakhstan dated 3 July 2014 №226-V of the Law of the Republic of Kazakhstan. – URL: https://adilet.zan.kz/eng/docs/K1400000226 (data obrashcheniya: 15.02.2025).

38 Ugolovnuyu otvetstvennost' za dipfejki mogut vvesti v Kazahstane. – URL: https://bluescreen.kz/ugholovnuiu-otvietstviennost-za-dipfieiki-moghut-vviesti-v-kazakhstanie/ (data obrashcheniya: 20.05.2025).

39 Kusainova I. Chislo ugolovnyh del za dipfejki rastet v Kazahstane. – URL: https://www.inbusiness.kz/ru/news/chislo-ugolovnyh-del-za-dipfejki-rastet-v-kazahstane (data obrashcheniya: 25.05.2025).

40 The Law of the Republic of Kazakhstan dated 21 May 2013 №94-V On Personal Data and their Protection. – URL: https://adilet.zan.kz/eng/docs/Z1300000094 (data obrashcheniya: 15.02.2025).

41 Mickaya E.V. Voprosy pravovogo protivodejstviya tekhnologii deepfake // Rossijsko-aziatskij pravovoj zhurnal. – 2025. – No1. – S. 50–58. – https://doi.org/10.14258/ralj(2025)1.8

Information about authors:

Smanova A. – corresponding author, Candidate of Law, Senior Lecturer, Kazakh National University named after al-Farabi, Al-Farabi Avenue, 71, 050040, Almaty, Kazakhstan

Muratova A. **–** Master in Chinese International Education, Mukhametzhan Tynyshbayev ALT University, Shevchenko st., 97, 050040, Kazakhstan

Zhumagulova Sh. **–** Candidate of Law, Senior Lecturer of the educational program «Jurisprudence», Korkyt Ata Kyzylorda University, Aiteke bi str., 29A, 120014, Kyzylorda, Kazakhstan

Сманова А.Б. **–** хат-хабар авторы, заң ғылымдарының кандидаты, аға оқытушысы, әл-Фараби атындағы Қазақ ұлттық университеті, әл-Фараби даңғылы, 71, 050040, Алматы, Қазақстан

Муратова А.Ж. **–** қытай халықаралық білімі магистрі, Мұхамеджан Тынышбаев атындағы ALT университеті, Шевченко көш., 97, 050040, Алматы, Қазақстан

Жумагулова Ш.Р. **–** заң ғылымдарының кандидаты, «Құқықтану» білім беру бағдарламасының аға оқытушысы, Қорқыт Ата атындағы Қызылорда университеті, Айтеке би көш., 29А, 120014, Қызылорда, Қазақстан

Сманова А.Б. – автор-корреспондент, кандидат юридических наук, старший преподаватель, Казахский национальный университет имени аль-Фараби, проспект Аль-Фараби, 71, 050040, Алматы, Казахстан

Муратова А.Ж. **–** магистр китайского международного образования, ALT Университет имени Мухамеджана Тынышпаева, ул. Шевченко, 97, 050040, Алматы, Казахстан

Жумагулова Ш.Р. – кандидат юридических наук, старший преподаватель образовательной программы «Юриспруденция», Кызылординский университет имени Коркыт Ата, ул. Айтеке би, 29А, 120014, г. Кызылорда, Казахстан

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https://creativecommons.org/licenses/by-nc/4.0/).