The Dimension and Limit of Criminal Sanction Thinking of Artificial Intelligence "Deep Counterfeiting"

Lin Ni *

Institute of Political Science and Law, Capital Normal University, Beijing,100048, China

* Corresponding Author Email: 1232903019@cnu.edu.cn

Abstract. The popularization of artificial intelligence "deep forgery" technology has spawned new crimes such as identity fraud and pornographic video forgery, which seriously impact social trust and citizens' personality rights, and challenge the criminal regulation system. Through case analysis and comparative law research, combined with technical deconstruction and judicial practice, this paper evaluates the interpretation space of crimes such as defamation and spreading obscene materials. The existing criminal law can lower the threshold of criminalization through judicial interpretation and cover most types of crimes. China should establish a regulatory framework that emphasizes "judicial optimization as the primary approach, with legislative prudence following suit." At the legislative level, priority should be given to improving the comprehensive protection of biometric information in the Personal Information Protection Law and the Cybersecurity Law, and considering the addition of minor offenses only when technological abuse severely impacts core legal interests. At the same time, it is necessary to strike a balance between technological innovation and risk prevention and control, and achieve collaborative governance through a combination of "technology + law," avoiding the suppression of the healthy development of artificial intelligence by overly harsh criminal policies. Ultimately, the goal is to achieve a dynamic equilibrium between the rule of law and technological ethics.

Keywords: Artificial intelligence, deep counterfeiting, criminalization, criminal law prevention, judicial response.

1. Introduction

Artificial intelligence "deep forgery" technology is rapidly popularized due to its open source and low threshold. Its technical nature is the abuse of artificial intelligence algorithms, and its behavioral nature is the abuse of personal biometric information, and technology abuse has given rise to new crimes such as identity fraud, infringement of personality rights (such as face-changing videos), spreading false information, and production of child pornography, which seriously threaten citizens' property rights, privacy rights, reputation rights, and the rights and interests of minors. Its high fidelity has shaken the foundation of social trust, and the traditional criminal law system is faced with the dilemma of "technological generational difference", so it is urgent to explore the regulatory path that adapts to the characteristics of technology. Domestic research focuses on the controversy over the expanded interpretation of crimes (such as defamation, child molestation), the strengthening of personal information protection, and the addition of identity theft. Foreign practice shows a distinction between "special legislation" (US) and "extension of data rights" (EU). However, they all face the difficulty of balancing technology abuse and freedom of speech. The existing research gaps focus on the lack of transnational crime cooperation mechanisms, the disconnection between technology identification standards and judicial evidence rules, the ambiguity of the supervision boundary of open source technology, and the lack of empirical analysis of the application effect of crimes.

This study takes the criminal risk caused by the abuse of "deep forgery" technology as the logical starting point, focuses on "technical characteristics - difficulties and challenges - perfection of formal regulation path", proposes the criminal sanction idea of "risk stratification + classified governance", defines the criminal law attribute of "deep forgery" through normative analysis, and promotes criminal legislation from passive response to active prevention. It provides theoretical support for the legal adaptability reform in the digital economy era.

For "deep forgery", an artificial intelligence technology that may be illegally used, it is of practical significance to explore the rational "dimension" of criminal law regulation of "deep forgery" technology, or to choose a relatively moderate legal governance countermeasure, and to clarify the limitations of criminal law in the face of emerging technologies, that is, the "limit".

2. The Characteristics of "Deep Forgery" Technology

The abuse of deep forgery technology lies in the illegal theft of personal biometric information. Biometric information covers physiological characteristics (such as fingerprints, iris, DNA) and behavioral characteristics (such as voice print, gait), which is unique, identifiable and irreversible. Artificial intelligence technology conducts deep learning and synthesis of biological characteristics through algorithms to generate fake audio and video, which directly infringes on individual identity rights and interests. Once such information is copied or abused, it will lead to systemic risks such as identity fraud and social credit collapse, and its correlation and irreversibility aggravate the proliferation effect of security vulnerabilities. Based on the special risk attributes of biometric information, it is urgent to build a full-chain regulation system through legislation, and strengthen the criminal sanctions and personal information protection mechanism in the case of technology abuse [1]. Deep learning technology based on generative adversarial networks (GANs) with high fidelity, low threshold and open source diffusion (such as the popularization of "one-click face change" tools). The neural network "illusion" principle and "deep forging" share the same underlying mechanism of generative AI, and its essence is to capture statistical rules in the data.

AI models may focus too much on local features (such as face textures) and ignore global logic (such as the number of fingers), resulting in the generation of content that defies common sense (such as a six-finger hand). This is similar to the phenomenon of "imagination" in the human brain.

The attacker takes advantage of the model's knowledge of data patterns by fine-tuning key parameters (such as facial feature vectors), allowing the model to miscombine the learned patterns (such as transplanting the features of A to B's face). At this time, the false content output of the model is more consistent with the statistical characteristics of the data, and it is difficult to identify with the naked eye. The technical homology of the two is reflected in the fact that as long as the AI model is better at "imagination" (generating realistic illusions), the easier it is to be manipulated to create highly simulated fake content. Current defense technologies identify anomalies by embedding digital watermarks and analyzing biometric consistency (such as blink rates), but fundamental solutions require refactoring AI's learning logic to preserve the basic constraints of the physical world while pursuing fidelity. This also reveals the central paradox of AI development: the difficult balance between creativity and authenticity.

3. The Existing Criminal Regulation Faces the "Limit" Challenge of the Emerging Technology of "Deep Counterfeiting"

3.1. The Type of Criminal Risk

“Deep fakes” are widely used in images, voice and video, posing a huge criminal risk. The intergenerational evolution of image "deep forgery" technology and the digital media ecology constitute a topological mapping, and its technical ontology has gone through the physical simulation stage which is limited by the silver salt development process, dependent on darkroom operation and microsculpture correction, with the PS tool is used to realize the digital editing stage of mass production, and the generative AI stage relies on GAN framework to complete the semantic reconstruction from latent space to 4K image. The "deep forgery" technology of speech relies on deep learning to realize multi-modal acoustic feature modeling, covering three dimensions of identity style, timbre and prosody. For example, the "Lyrebird" system of the University of Montreal uses Bi-LSTM to model emotion vectors and combines style transfer algorithm to simulate pragmatic features [2]. The prosody model optimized by counter-training has a MOS score of 4.2, with cross-language

generalization ability.The "deep forgery" technology of video is based on the deep neural network architecture, and multi-layer nodes are used to represent and learn the multi-modal biometric features of the target individual, and nonlinear mapping models are constructed by using GAN and its variants (such as StyleGAN and CycleGAN) to achieve cross-domain synthesis of facial expressions, lip movements and facial topologies.

3.2. Infringement of Property Rights

The infringement of property rights by artificial intelligence "deep counterfeiting" technology follows the logical chain of "technology enablement - trust subversion - right alienation". It breaks through the identity verification mechanism through high-precision simulation of biometrics (such as voice print and micro-expression), and uses human cognitive inertia of audio-visual evidence to reconstruct the "real" decision-making scene, making victims fall into the wrong expression of meaning. Taking the case of forged executive instructions in Hong Kong in 2024 as an example, deep counterfeiting not only copies biometric characteristics, but also trains simulated decision-making habits through dialogue logic, so that the fraudulent behavior has "authorized rationality", and directly cuts off the meaning connection between the property disposition behavior and the real right holder. At the legal level, the nature of technology dispels the linear causal relationship of traditional tort "act - fault - damage". The George Carlin case in the United States broke through the traditional framework of personality rights, incorporated AI-generated content into the object of "digital property rights", and reconstructed the rules of liability allocation by expanding the extension of "actionable acts".

3.3. Infringement of Personality Rights

The infringement of personality rights by deep forgery technology presents a progressive logic of "technology penetration - trust alienation - rights depletion".

Its double harm mechanism could be seen through some classic cases: the direct type such as the Hong Kong executive counterfeiting case, through the simulation of biometric characteristics and decision-making habits to break through the corporate financial authorization system, tampering with the intention of personality disposal; Indirectly, such as the case of female journalists in India, fake pornographic content leads to lower social evaluation and thus occupational income loss. Its technical essence lies in deconstructing the two fundamental property rights of "identity authenticity" and "meaning authenticity".

3.4. Infringement of National Security and Public order

"Deep forgery" technology systematically undermines the national governance order by deconstructing social trust anchors and legal and factual basis.

Its core logic is as follows: first, it subverts the authenticity of news with highly simulated audiovisual data (for example, Pew report shows that 63% of Americans change their information acquisition habits), and the algorithm recommendation mechanism accelerates the spread of false information, forming a transmission chain of "cognitive pollution - trust collapse - order disorder"; Secondly, the falsification of evidence directly impacts the judicial evidence chain (such as the "circuit Prosecution Team" tampering with the monitoring plot), breaking through the "evidence authenticity" this cornerstone of the rule of law; Thirdly, in April 2018, a "deep fake" video featuring former US President Barack Obama appeared on the US social networking service platform Twitter. In the picture, he is sitting in his office and forcefully saying "Trump is a total idiot", and in the later part of the video, the picture is split into two, the left is still Obama speaking, and the right is a person with the same expression and the same words as Obama. The video, intended to warn people against "deep forgery" technology, has been viewed more than 2 million times and received more than 50,000 likes in just half a day. Once the false video distorting the political remarks of national leaders is widely spread on the Internet, it is bound to tarnish the image of national leaders and even trigger

geopolitical risks [3]. The data shows that the content of deep forgery has increased by more than 330% year on year, exposing the lagging regulation of the law on "technology-based social harm".

3.5. Dilemmas and Challenges of Existing Criminal Regulation

The current legal framework still has limitations and technical obstacles when dealing with the crime of artificial intelligence "deep forgery".

4. Charge Application Dispute

Ambiguity of Criminal Standards. Existing crimes (such as libel and fraud) need to be interpreted to cover new behaviors, but there is a problem of ambiguity of criminal standards (such as "serious circumstances" determination), and the ambiguity of criminal standards of artificial intelligence deep counterfeiting stems from the dilemma of defining technical neutrality and the degree of legal interest infringement. The law needs to balance technical ethics and social harm, but the concept of "forgery" in the current criminal law is difficult to cover the multidimensional impact of deep forgery on personality rights and public order. Taking the case of "deep counterfeiting," Biden's voice intervention in the primary election in the United States as an example, the producer was acquitted due to the freedom of speech defense, exposing the separation of subjective malice and objective harm. The core contradiction of the vagueness of the criminalization standard lies in the lack of a quantitative standard of the boundary between the freedom of technical creation and the protection of legal interests, and the logical conflict of the identification of the behavior crime and the result crime.

Breaking through the Constitutive Requirements of Traditional Crimes. Taking the regulation of child pornography as an example, artificial intelligence deep forgery of child pornography should break through the traditional crime of child molestation: extend "children's sexual autonomy" to "children's figurative sexual dignity", and recognize the collective legal interest infringement of virtual images on the symbolic exploitation of children's groups. By analogy with Section 2256 of the US Child Protection Act, which includes deeply falsified data "identifiable as a child" in the category of "child pornography", the responsibility is pursued in the mode of behavioral crime, without the actual harm result, and the objective danger of "subjective cognition + content in line with the sexualized characteristics of children" is emphasized [4].

4.1. Technical Challenges

The technical challenges of "deep forgery" technology screening mainly include: Evidence is difficult to fix, the identification technology of AI-generated content is immature, and judicial forensics relies on technical detection tools (such as blurred boundaries and hand details defects); It is difficult to define the responsibility of the platform, and the conflict between the censorship obligation of the communication platform and the privacy of the user (such as the failure to label the composite content).

4.2. Legal Lag

Existing regulations (such as the Interim Measures for the Management of Generative Artificial Intelligence Services) lack targeted criminal provisions and vague punishment standards, which are mainly reflected in: The current legal interest protection scope does not extend to the collective rights and interests of virtual images. The traditional crime of child molestation focuses on the actual child victimization, while the deep forged child pornography only needs the synthetic data "identifiable as children" to complete the symbolic exploitation of children's collective dignity. Moreover, technical neutrality leads to the difficulty in determining subjective malice. For example, the US Child Protection Act takes "manufacturing intent + sexualized characteristics of children" as the criminalization standard, while China's criminal law still relies on the actual harm result and cannot effectively regulate potential risks. The problem of crime competition is prominent, that is, the same act may involve the dissemination of obscene materials, infringement of personal information and

other charges, but the lack of a unified quantitative standard, resulting in judicial discretion out of focus.

5. The Construction and Improvement of the Criminal Regulation Path "Dimension" of "Deep Forgery" Technology

5.1. Promoting Artificial Intelligence Deep Forgery Detection Technology

5.1.1. Technology Research and Development and Standard Construction, Algorithm Upgrade and Multi-Mode Detection

Based on the technical characteristics of generative adversarial network (GAN), the detection algorithm needs to break through the traditional single-dimensional feature analysis and turn to multi-modal data fusion analysis such as facial micro-expression, voice print continuity and limb movement timing. For example, the whole body forgery detection model developed by the University of California, Berkeley in the United States can identify forged videos through the abnormal trajectory of joint motion, and the error rate is reduced by 40% compared with the traditional method [5].

5.1.2. Open Source Database and Certification Standards

Establish a national deep forgery sample database, collect typical forgery feature data, and formulate performance evaluation standards for detection tools. For example, the EU GDPR requires detection tools to be certified by ENISA (European Union Cyber Security Agency) to ensure a false positive rate of less than 5% [6].

5.1.3. Identification of Evidence Qualifications

According to Article 93 of "Several Provisions of the Supreme People's Court on Evidence in Civil Proceedings", the test report must meet the dual requirements of "professional organizations + standard procedures". It is suggested to set up special qualifications for forensic AI testing, such as referring to the United States "Deep Forgery Responsibility Act", requiring testing institutions to have ISO/IEC 17025 certification, and use the algorithm tool registered by the National Cyberspace Administration [7].

5.1.4. Hierarchical Rule of Proof Force

To ensure the basic proof force, the test report should be marked with the value of forgery probability (such as >90% for highly credible), and attached with the hash value of the original data for verification.

As for the reinforcement rule, a single technical evidence should be combined with victim statements and transmission path analysis to form an evidence chain. For example, in Baotou's "AI face exchange fraud case of 4.3 million", the court simultaneously reviewed the transfer records and IP traceability data when accepting the credit test report [8].

5.2. System Coordination and Risk Prevention and Control

5.2.1. Platform Responsibility Pre-positioning

Referring to Article 11 of the "Regulations on the Management of Network Audio and Video Information Services", social platforms are required to embed real-time detection plug-ins, and the content that fails to pass the detection is forced to mark "composite". For reference, the United States "Deep Counterfeiting Liability Act" stipulates that platforms that fail to fulfill their marking obligations shall bear joint and several liability.

5.2.2. Ethical Boundaries of Technology

Relevant laws and regulations require technology providers to embed ethical design (such as open source tools to limit abusive features); The application of detection technology should follow the principle of proportionality to avoid excessive collection of biometric data. The Swedish data

regulator can learn from the penalty case of the face recognition attendance system, requiring the detection tool to strip the personally identifiable information and only output the binary identification conclusion.

5.3. Empirical Effect and Correction Mechanism

The dynamic evaluation model is established to analyze the influence of detecting technology's miscarriage rate on judicial justice regularly. For example, the German Federal Court introduced the "technical evidence objection hearing" procedure in 2022, allowing the parties to apply for a third-party agency to reexamine, and the error rate exceeds the threshold to start the evidence exclusion rule.

5.4. Direction of Law Improvement

The essence of technology abuse is the theft of identity symbols. Legislation is needed to fill the "access - use" fault, technical labeling reduces the difficulty of proof, platform filtering and blocking transmission, and form a full chain of "manufacturing - circulation - harm".

5.5. Adding Charges and Platform Liability

Adding Special Charges. Introduce the crime of "identity theft" or extend the crime of "infringement of citizens' personal information" to the misuse of biometric information. Governance of artificial intelligence "deep counterfeiting" needs to build a "legislative stratification + technology embedding + responsibility chain traceability" system:

First, legislatures need to implement a hierarchical mechanism of legislation. Amend Article 253 of the Criminal Code to clarify separate subcategories of "biometric information" (including face, voice print, gait, etc.), Add "crime of illegal use" to regulate abuse after legal acquisition (such as face-changing synthesis); The new crime of "identity theft" is added at the same time, and the use of biometric characteristics to impersonate others to commit fraud and other acts is increased penalties (refer to the Texas "Forged Video Influence Election Act").

Then, the judicial authorities need to apply for judicial expertise in advance. Such as the establishment of a national biometric information forensic authentication center, the use of multi-modal cross-verification (such as the Tsinghua University DeepFakeDetect system) and the detection report as the core evidence of conviction.

Next, the technical authorities need to establish the technical criteria for criminalization. Such as forcing deep forgery tools to embed digital watermarks and metadata labels (such as ISO/IEC 30107 standards), judicial authorities can trace the algorithm hash value, and those who do not mark it are regarded as "malicious forgery" directly.

Finally, the platform is required to bear joint liability. According to the Regulations on the Management of Network Audio and Video Information Services, the platform is required to deploy AI detection interfaces (error rate $\leq 3\%$), and failure to intercept high-risk forged content is held accountable as an aid.

Platform Responsibility Compaction. First of all, we need to establish an advanced compensation mechanism. According to Article 47 of the Network Security Law, if the platform fails to perform the AI detection obligation and causes damage to the user's property or reputation, it shall start the "presumption of fault" to pay compensation in advance (such as 50% advance payment according to the actual loss), and then seek compensation from the infringer [9]. For example, the Deep Counterfeiting Act of Virginia in the United States stipulates that the platform shall bear a single fine of $100,000 if it fails to mark the forged content.

Then the dynamic review criteria need to be refined. The platform must deploy a multi-modal detection system (such as Microsoft Video Authenticator) to scan uploaded content in real time, and the misjudgment rate must be lower than 2.5% of the EU GDPR requirements. The detection model must be recorded through the algorithm of the National Cyberspace Administration, and the countersample database should be updated every quarter.

Finally, platforms need to strengthen the embedment of user education. Force platforms to add "deep forgery risk alert floating Windows" to content publishing pages and provide one-click AI detection portals (such as Adobe's Content Credential Tool). Anti-fraud case teaching videos are targeted to high-risk users (such as the elderly and low-education groups), with a monthly coverage rate of more than 95%.

In short, it is to force technological upgrading for economic compensation, dynamic review to block the source of transmission, education embedment to enhance users' autonomous defense capabilities, and form a "prevention - interception - relief" closed-loop governance.

5.6. Gradient Construction and Implementation of Criminal Sanctions System

Improving the Range and Accuracy of Sentencing. The punishment of malicious production and dissemination is increased, and the compliance exemption clause is set up for technology development to achieve the balance of crime, responsibility and punishment in essence.

First of all, the technology level weighting is carried out. Based on the technical complexity of criminal tools (e.g. GAN generative adversarial network, LSTM timing model) to divide the sentencing base, each upgrade of a technical level (according to IEEE standards) increases the base sentence by 20%. For example, the sentence for using the open-source face changer tool is 3 years, and the self-developed adaptive algorithm is increased to 5 years.

Secondly, the cost of restoration should be offset. The defendant is required to bear the cost of AI detection removal (such as paying 500 yuan for every 10,000 transmissions eliminated), and the full performance can reduce the sentence by 30%. The Ministry of Justice established a national deep counterfeiting restoration Fund to enforce the recovery in accordance with Article 1182 of the Civil Code.

Promoting cross-border collaboration (take the joint fight against the transnational spread of child pornography as an example). Actively promote cross-border collaboration, that is, data sharing breaks through technical anonymity, algorithm mutual recognition reduces judicial barriers, a circuit breaker mechanism accurately strikes crime ecology, and forms a global co-governance paradigm of technological crimes.

First, a global database of characteristics could be established. With reference to Interpol ICSE (Child Sexual Exploitation Database), establish a multi-modal deep forged child pornography sample signature database (including facial hash, voiceprint map, GAN generation track), National law enforcement agencies are updated synchronously in real time to enable blockchain storage to ensure that the chain of evidence is valid across borders [10].

Second, a mutual recognition agreement for detection algorithms can be signed. Based on ISO/IEC 23894 standard, multinational certification of mainstream detection tools (such as Microsoft PhotoDNA, Google Content Safety API), the test results are included in the mutual recognition framework of electronic evidence of the Hague Convention on Forensics to avoid duplicate identification.

Third, a joint law enforcement circuit breaker is needed. Deploy the AI honey pot system on the active nodes of the dark net, automatically track the transnational IP jump path, synchronously freeze 27 virtual currency money laundering channels around the world when triggering the "circuit breaker" (refer to the 16th recommendation of the International Anti-Money Laundering FATF), and launch the "Guardian" transnational joint arrest action through the Europol platform.

6. Conclusion

In view of the technical risks hidden by "deep forgery", a new thing, the various related parties can establish different criminal law regulatory positions, which depend on our positioning of "deep forgery" technology itself. Existing research shows that the abuse of deep forgery technology poses a severe challenge to personal biometric information, privacy and data security, and impacts the traditional criminal law system. It emphasizes the need to balance technological innovation and legal

interest protection. Although the existing criminal law can deal with some risks by adjusting the criminalization standard, there are structural deficiencies in the regulation of data utilization behaviors. It is necessary to add illegal data analysis crimes and data manipulation crimes, and improve the constitutive elements of the crime of refusing to perform the information network security management obligations. Legislative branch need to strengthen the hierarchical protection of biometric information through criminal law, and supplement the "legal acquisition + illegal use" of the standardized blank identity theft penalty, data classification and classification of governance and the proposed legal benefit protection model, which provides theoretical support for the construction of a criminal law framework to adapt to the digital age. At the same time, in judicial practice, attention should be paid to the quantitative impact of technical characteristics on the degree of infringement of legal interests, and mechanical application of communication standards should be avoided. The follow-up research should focus on the elaboration of the constituent elements of new crimes, transnational legal coordination and technical governance coordination. How to quantify the technical elements of "serious circumstances", develop efficient detection tools to assist judicial determination, and explore the interface mechanism between corporate compliance and criminal responsibility are urgent directions to break through.

References

[1] ZHAO X.: Criminal Law Evaluation of Artificial Intelligence "Deepfake" Technology. Journal of Suzhou Education Institute 24 (4), 99 - 110 (2021).

[2] FAN Y., YU Y.: The Criminal Law Regulation of "Deepfake" Technology and its Products in Network Communication. East China University of Political Science and Law, Shanghai (2021).

[3] TAN B., CHANG H.: Seeing is believing? "Deepfakes." Learn about it. People's Daily Online (2017), http://world.people.com.cn/n1/2017/0103/c1002-28995895.html, last accessed 2022 – 11 - 15.

[4] U.S. Department of Justice: Citizen's Guide to U.S. Federal Law on Child Pornography. 18 U.S.C. § 2256 (2020).

[5] LI H.: On the Criminal Responsibility about the Abuse of Personal Biometric Information - Taking Artificial Intelligence "Deepfake" as an Example. Tribune of Political Science and Law 38(4), 1-15 (2020).

[6] WANG Y.: Criminal Regulation on Infringement of Citizen's Right of Personality by "Deepfake" Technology: Case Study of Fake Face-changing Pornographic. Journal of Henan Police College 32 (3), 99 - 110 (2023).

[7] JIANG Y.: On the Dimension and Limitation in Criminal Regulation of AI "Deepfake". Journal of Nanjing University (Philosophy, Humanities and Social Sciences) 58 (9), 1 - 12 (2021).

[8] YU S.: Study on the Prevention and Control Countermeasures of "AI Changing Face" Impersonating Identity Fraud Crime. Law Fair 15 (3), 1 - 5 (2024).

[9] Standing Committee of the National People's Congress: Cybersecurity Law of the People's Republic of China. (2016), http://www.npc.gov.cn/englishnpc/lawsofthePRC/202103/82a8f60705f445e5b6cdd01c4e2136b2.shtml, last accessed 2024 - 05 - 20.

[10] YANG J.: The Regulatory Dilemma and Practical Way Out of the Crime of Deepfake Child Pornography. Journal of Hainan University (Humanities & Social Sciences) 42 (5), 1 - 10 (2024).