ISSN: 2710-0170

New Trends and Privacy Protection Challenges of Cybercrime in the Era of Artificial Intelligence

Zhixi Chen

School of Academy for Applied Policy Studies and Education Futures, The Education University of HongKong, HongKong, China

Abstract

The rapid development of artificial intelligence technology is reshaping the form of the digital economy, while also providing precise, large-scale, and covert technological empowerment for cybercrime, giving rise to new forms of crime such as AI face swapping and counterfeiting AI models to spread Trojans, posing a serious threat to personal privacy and data security. This article is based on authoritative sources such as the "Clean Net 2025" special action case of the Ministry of Public Security and Check Point's "AI Security Report". It systematically analyzes the technical application path of artificial intelligence in cybercrime, summarizes new trends such as the popularization of criminal tools (such as low threshold face swapping software), the precision of attack methods (such as customized fraud scripts), and the collaboration of criminal modes (such as technology logistics division of labor assembly lines); Deeply explore the current situation of expanding privacy leakage scenarios under new types of cybercrime, and analyze the multiple challenges of privacy protection in legal adaptation, technological defense, regulatory governance, and other aspects; Finally, a multidimensional response framework of "technical defense+ legal regulation+ regulatory coordination+ public literacy" was proposed. Research has shown that the double-edged sword effect of artificial intelligence has profoundly changed the pattern of cybercrime, and the privacy protection system needs to transform from passive defense to active prevention and control, achieving a dynamic balance between technological innovation and security through multi-party collaborative governance. This article provides practical reference and theoretical support for addressing network security risks and improving privacy protection mechanisms in the AI era.

Keywords

Artificial intelligence, cybercrime, privacy protection, polymorphic phishing, collaborative governance.

1. Introduction

With the popularization of generative AI and deep learning technologies, the digital space attack and defense situation has undergone fundamental changes. Previously, cybercrime that relied on professional programming skills has now achieved a breakthrough in "populism" through AI tools - the FraudGPT tool circulating on the dark web, which starts at $200 per month, allows non-technical individuals to generate phishing emails, malicious code, and counterfeit websites with just one click; Legitimate tools such as ChatGPT have also been misused to quickly generate customized scam content, significantly reducing the threshold for attacks [1]. According to KnowBe4's "Q2 2025 Simulated Phishing Report," AI phishing attacks disguised as internal emails have a click through rate of up to 98.4%, with over 80% using "domain spoofing" technology to forge trusted sources that traditional protection systems find difficult to identify. The abuse of AI technology not only increases the success rate of attacks, but also expands the risk of privacy breaches: from biometric infringement caused by AI face

ISSN: 2710-0170

swapping and voice cloning, to trade secret leaks caused by polymorphic phishing, to AI collaborative ransomware stealing medical and financial data, the harm far exceeds traditional crimes. The "Clear and Rectify AI Technology Abuse" special action deployed by the Cyberspace Administration of China in 2025 clearly states that AI abuse has formed a complete black industry chain covering illegal tool development, tutorial sales, and precise attacks. Although China's Personal Information Protection Law and Cybersecurity Law have laid the legal foundation, there are still problems such as vague characterization and lagging defense in the face of new types of crimes. How to grasp the evolution of crime and solve the dilemma of privacy protection has become an urgent issue for the development of the digital economy. Based on this, this article combines real cases and industry data to conduct research, providing reference for building a security governance system in the AI era.

2. The technological foundation and tool evolution of AI empowering cybercrime

The empowerment of artificial intelligence for cybercrime stems from the collaboration of generative models, big data analysis, and automation tools, driving the transformation of crime from a "manual workshop" to an "intelligent factory". The technological foundation focuses on natural language generation, computer vision synthesis, and automated attack script development, and the commercialization of criminal tools further lowers the threshold. The abuse of generative language models is a prominent feature. FraudGPT, WormGPT, and other malicious tools based on GPT-3 are subscribed and circulated on the dark web on a monthly basis. The core functions include customized phishing emails, fraudulent documents, and lightweight malware. Users do not need programming skills and can prepare for attacks through natural language commands. This type of email can be combined with public information on corporate websites and workplace platforms, imitating the tone of the target organization and mentioning real projects and personnel, which is far more deceptive than traditional content. The Public Internet Anti phishing Working Group (APWG) pointed out that AI has significantly shortened the preparation time for harpoon phishing and achieved large-scale replication of precision fraud. Computer vision and speech synthesis technology have given rise to identity forgery crimes. The Central Cyberspace Administration's "Clear and Bright" special action report shows that illegal AI products can "change faces with one click" and "clone voices", illegally clone biometric features, and use them to impersonate relatives, friends, and corporate executives for fraud. This technology not only reduces the difficulty of forgery, but also has concealment. Attackers can generate fake executive avatars and voice messages, and cooperate with forged instructions to carry out fraud, making it difficult for victims to distinguish authenticity. In a recruitment scenario of a certain enterprise, attackers use AI to synthesize candidate information and implant malicious scripts, infiltrate the supply chain through resume delivery, resulting in the leakage of research and development data, highlighting the chain harm of biometric forgery [2]. Automated attack tools further amplify the scale of crime. The AI driven platform can automate the construction of phishing websites, generation of malicious code, and screening of attack targets. Attackers can use AI to generate code and quickly build counterfeit pages that are highly similar to the official website, hosted on legitimate cloud platforms to obtain SSL certificates, and use the "security lock" symbol to reduce user vigilance; At the same time, by analyzing massive amounts of data through AI, individuals and enterprises with weak protection can be accurately identified, achieving "targeted clearance". This "tool automation+target precision" mode has shifted network attacks from "casting a wide net" to efficient and precise attacks, significantly improving the success rate.

ISSN: 2710-0170

3. The core new trend of cybercrime in the era of artificial intelligence

The deep penetration of artificial intelligence technology has led to the emergence of four new characteristics in cybercrime: the democratization of criminal thresholds, the polymorphism of attack methods, the synergy of criminal chains, and the clustering of infringement areas, posing a comprehensive threat to privacy and security. The democratization of the criminal threshold is the most significant. Traditional cybercrime requires professional programming and infiltration skills, and the popularization of AI tools is changing this situation. The dark web is filled with "foolproof" AI crime tools, and attackers can complete the entire process of phishing email generation, malicious development, and target screening by paying for subscriptions or purchasing tutorials. WormGPT costs 60 euros per month and can provide code and Trojan functionality. Even those without technical expertise can carry out complex attacks. Reed Flute of the public Internet anti phishing group pointed out that AI is promoting the "civilians" of cybercrime. Attacks that used to require professional skills can now be realized through simple instructions, directly expanding the criminal community. The polymorphism of attack methods is the key to breaking through defense. Traditional fixed feature protection is difficult to deal with AI dynamic attacks: Attackers use AI to generate a large number of phishing email variants, with only minor differences, forming "polymorphic phishing"; It also uses interference detection systems such as zero width characters. KnowBe4 revealed that over 70% of phishing attacks in 2024 have polymorphic features, bypassing blacklist detection through dynamic camouflage, and shifting attacks from "fixed templates" to "real-time evolution", posing a risk of traditional protection failure. The synergy between cybercrime and ransomware has significantly increased. AI reduces phishing costs and becomes a carrier for ransomware dissemination. In the fourth quarter of 2024, there was a significant increase in ransomware spread through AI phishing emails, with attackers adopting a strategy of "polymorphic phishing+triple encryption+supply chain poisoning". In a medical institution case, the attacker used AI to generate fake firmware packages, sent phishing emails to nurses, encrypted the system within 30 minutes, and stole medical records, resulting in a double blow of "data theft+system paralysis" [3]. Focus on high-value targets in the field of infringement. Most AI phishing attacks target engineering and technical personnel, using resume poisoning and interview traps for implementation; Enterprise executives are the core targets of spear phishing, and attackers use public information to generate customized emails to induce leaks. Recruitment season attacks are rampant, with fake "personality test links" and Trojan horses disguised as onboarding instructions becoming a means of stealing secrets. In a certain technology company, an attacker sent an AI "quarterly summary template" to the R&D email, and interns downloaded it and implanted a backdoor into the R&D environment, highlighting the precise harm.

4. Scenario expansion and risk escalation of privacy breaches under new types of cybercrime

The abuse of artificial intelligence technology has pushed privacy breaches from traditional data theft to multi-dimensional expansion, with biometric information, supply chain data, and medical and financial privacy becoming the focus of attacks, showing an upgraded trend of rapid spread, wide scope, and deep harm. Biometric information is a high-risk area for privacy breaches. The illegal application of AI face swapping and voice cloning technology poses serious risks to facial, speech, and other biometric features. The special action of the Cyberspace Administration of China pointed out that a large number of unauthorized AI products have cloned other people's voiceprint and facial information for fraud or marketing purposes. This type of information cannot be changed like a password, which poses a long-term risk of identity theft for victims. In 2023, a Hong Kong company was once subjected to voice fraud by AI cloning

ISSN: 2710-0170

of its parent company's CFO, resulting in huge financial losses and highlighting the direct harm. AI can also optimize public facial photos and voice clips to generate highly realistic data, amplifying the risk of leakage. Supply chain penetration leads to horizontal expansion of leakage scenarios. Attackers no longer directly attack targets, but instead use AI to infiltrate third-party weak links. A large number of AI network attacks rely on third-party platform trust endorsements and use partner enterprise email addresses to launch attacks, breaking through the protection boundaries of a single organization. In 2025, cybersecurity giants such as Palo Alto Networks and Cloudflare were invaded by third-party application Drift, resulting in Salesforce data breaches that affected hundreds of enterprises. This type of attack spreads privacy breaches from a single point to the industry chain, making the aftermath extremely difficult. The risks in sensitive fields such as healthcare and finance are prominent. Some registered AI products provide medical Q&A services in violation of regulations, resulting in the leakage of medical record information; AI phishing in the financial field accurately targets account information and carries out fraud by forging official websites and apps. Attackers often generate subdomains similar to the official website, and use compliant phishing emails to bypass defense systems and steal sensitive user data. Such leaks often cause irreversible damage [4]. The concealment and persistence of privacy breaches have significantly increased. The combination of AI and file free attack technology enables malicious code to run without writing to disk, making it difficult for traditional antivirus software to detect; The implanted persistent backdoor can steal data for a long time. When a medical institution was subjected to a ransomware attack, the attacker simultaneously implanted a latent module and continued to steal medical records for three months before being discovered, causing the victim to suffer long-term damage without knowledge.

5. Multidimensional adaptation challenges faced by privacy protection systems

The multidimensional adaptation challenges faced by the privacy protection system in the face of AI driven new types of cybercrime and privacy risks, the current privacy protection system is facing challenges in the fields of law, technology, regulation, ethics, and public literacy, and traditional models are difficult to cope with dynamic threats. There is a lag and insufficient adaptation of legal regulations. Although the Personal Information Protection Law and the Cybersecurity Law clearly define the protection of personal information, there is a lack of detailed rules for determining and assigning responsibility for infringement of AI generated content, such as gaps in the infringement subjects of AI face swapping and voice cloning; The legality of generative AI training data is highly controversial, and it is difficult to determine whether online crawling of personal information for training is illegal. The existing criminal charges in the Criminal Law are difficult to adapt to AI assisted crime characterization, and the legal responsibility definition for algorithm recommendation and AI anti detection is vague, making it difficult to sanction some infringements. Technical defense has gaps in storage and collaboration. Traditional passive defense is difficult to resist AI active attacks: feature library blacklist protection is ineffective against polymorphic phishing and dynamic malicious code; Data encryption is difficult to resist AI brute force cracking. Although privacy enhancing technologies (differential privacy, federated learning) have potential, they have poor compatibility and significant performance loss, making them difficult to promote. The gap between attack and defense iterations is large, attackers quickly use new AI technologies, enterprise protection updates lag behind, and small and medium-sized enterprises have difficulty deploying AI detection due to cost, amplifying vulnerabilities [5]. Regulatory governance faces challenges in cross domain collaboration and technology identification. AI cybercrime is cross regional and highly anonymous. Attackers use the dark web and cross-

ISSN: 2710-0170

border servers to commit crimes, and legal differences and law enforcement barriers hinder cooperation; The rapid dissemination of AI content leads to the rapid spread of infringement, and regulatory response lags behind. Moreover, AI generated content detection is difficult, deep synthesis of implicit identifiers is not widely used, and some social platforms do not strictly control AI services accessed through APIs, leaving room for infringement. The issue of ethical balance and public literacy is highlighted. The development of AI requires massive data training, which is in tension with privacy protection. Overprotection hinders innovation, and the use of misconduct increases risks. In addition, there is a lack of ethical guidelines to guide balance. Some companies excessively collect sensitive information, resulting in a "heavy use, light protection" pattern. The public's awareness of new risks is insufficient, and traditional protective habits are still being used. The literacy of enterprise employees needs to be improved, and there is a lack of AI phishing and deep forgery exercises in training.

6. Collaborative governance framework for addressing AI cybercrime and privacy risks

To address the challenges of cybercrime and privacy protection in the AI era, it is necessary to establish a multidimensional collaborative governance framework of "technological defense, legal regulation, regulatory innovation, and public literacy", and achieve a transformation from passive response to active prevention and control, and from single governance to collaborative governance. Technological defense needs to be upgraded to AI driven proactive prevention and control. Enterprises should deploy dynamic detection security systems and use AI to combat AI crimes: identify AI generated content features through behavior analysis, build a zero trust architecture, and implement multi factor authentication for privileged accounts; Promote federated learning and differential privacy technology, popularize end-to-end encryption, and ensure the security of the entire data process; Establish a security self-assessment mechanism and regularly investigate AI privacy risks. Legal regulations need to be improved and dynamically adapted. At the legislative level, refine AI privacy protection rules, clarify the legal attributes and infringement responsibilities of AI generated content, revise the Personal Information Protection Law, standardize the legality of AI training data, and establish a corpus review mechanism; Strengthen guidance on new types of infringement cases at the judicial level, clarify the conviction standards for AI face swapping and voice cloning; Promote international legislative coordination, draw on the experience of the European Union's GDPR, and establish cross-border judicial assistance mechanisms. The regulatory mechanism needs to achieve precision and full chain coverage. Implement the AI product filing system and supervise the classification and grading of generative AI; Establish an AI generated content identification system, requiring the addition of explicit or implicit identification; Carry out special rectification and clean up illegal AI products and tools; Build a national unified monitoring platform and use big data and AI to warn of risks; Strengthen the linkage between departments such as cyberspace, public security, and industry and information technology to form a regulatory synergy [6]. Industry self-discipline and public literacy need to be strengthened simultaneously. Industry organizations formulate self regulatory norms and technical standards for AI privacy protection; Enterprises embed privacy protection into the entire AI process, implement the principle of "privacy design", and conduct practical exercises for AI phishing and deep forgery; Popularize AI crime prevention techniques through media and community lectures, with a focus on enhancing the awareness of elderly and youth protection, promoting privacy protection tools, simplifying the process of infringement complaints, and creating a protective atmosphere with public participation.

ISSN: 2710-0170

7. Conclusion

While artificial intelligence promotes social progress, it also reshapes the form of cybercrime, presenting features such as tool populism, precise attacks, polymorphic means, and deep harm, posing severe challenges to personal privacy and data security. The abuse of generative AI reduces the threshold for criminal activities, leading to frequent occurrences of new types of crimes such as polymorphic phishing, deep forgery, and supply chain collaborative attacks. Privacy leakage scenarios continue to expand, with biometric information and sensitive domain data becoming the focus of attacks. The concealment and persistence of privacy risks have significantly increased. The current privacy protection system is facing multiple challenges: lagging legal regulations and difficulty in covering new types of privacy infringement; Technological defense is outdated, and traditional methods are difficult to resist AI proactive attacks; Regulatory challenges include cross domain collaboration and technical identification, making it difficult to achieve full chain control; Insufficient public digital literacy can easily lead to privacy breaches. The traditional single dimensional protection mode is no longer suitable for the needs of the AI era, and a multi-party, multi-dimensional collaborative governance system needs to be built. We must adhere to the principles of balancing technical defense and legal regulation, coordinating regulatory governance with industry self-discipline, and balancing risk prevention and technological innovation. By deploying AI proactive defense systems, improving privacy protection regulations, innovating precise supervision mechanisms, strengthening corporate responsibility, and enhancing public literacy, a comprehensive protection network is formed; At the same time, promote global governance collaboration, establish cross-border AI crime judicial assistance mechanisms and technical standards, and address global privacy threats. Privacy protection in the AI era is a guarantee for sustainable technological innovation. Only by building a sound collaborative governance system can we curb the abuse of AI, prevent criminal risks, and achieve a dynamic balance between technological innovation and privacy protection. In the future, the government, enterprises, social organizations, and the public need to continue to make efforts to optimize governance strategies and create a safe environment for the healthy development of the digital economy.

References

[1] Kulothungan V, Gupta D, Kandel L N. Democratizing Cybercrime: Risks and Countermeasures of AI-Enabled Attacks[J].

[2] Sharma A, Kejriwal D, Pakina A K. Generative AI in the deep internet: Treat and counter measures for the next generation resilience[J]. International Journal of Artificial Intelligence and Data Research, 2023, 14(1).

[3] Alessandro R, Giulia B. AI-Enhanced Cybersecurity Proactive Measures against Ransomware and Emerging Threats[J]. Innovative: International Multi-disciplinary Journal of Applied Technology, 2024, 2(11): 77-92.

[4] Lamina O A, Ayuba W A, Adebiyi O E, et al. Ai-Powered Phishing Detection And Prevention[J]. Path of Science, 2024, 10(12): 4001-4010.

[5] Xu R, Baracaldo N, Joshi J. Privacy-preserving machine learning: Methods, challenges and directions[J]. arXiv preprint arXiv:2108.04417, 2021.

[6] Omopariola M. AI-ENHANCED THREAT DETECTION FOR NATIONAL-SCALE CLOUD NETWORKS: FRAMEWORKS, APPLICATIONS, AND CASE STUDIES[J]. 2017.