Legal Risks of the Application of Artificial Intelligence Anti-Trafficking System and Its Regulatory Paths

Gao Wanwen[1]*, Chen Yaxin[2], Huang Yaling[3]

[1] *Law Department, Minnan Normal University, China; 1736713659@qq.com

[2] Law Department, Minnan Normal University, China; 2465964070@qq.com

[3]Law Department, Minnan Normal University, China; 709362304@qq.com

*Corresponding Author: 13559152404@163.com

DOI: https://doi.org/10.30212/JITI.202503.009

Submitted: Mar. 31, 2025 Accepted: May. 26, 2025

ABSTRACT

The application of artificial intelligence (AI) in the field of anti-trafficking presents both technological empowerment and social risks. Currently, its use is becoming increasingly widespread, but the legal risk it brings including data security, algorithmic bias, and liability attribution, are gradually becoming more prominent, and thus, requires in-depth research and effective regulation. This paper aims to comprehensively analyze the legal risk associated with AI anti-trafficking systems and explore the construction of a scientifically sound regulatory framework to balance technological development with rights protection. This paper, through empirical analysis and normative research, explains the application of artificial intelligence anti-trafficking systems in intelligence assessment of trafficking-related information, intelligent collection of evidence related to trafficking, and intelligent interrogation of suspects involved in trafficking. It reveals new legal risks such as data security, algorithmic discrimination, and liability attribution that arise from these applications. Moreover, based on the framework of technology governance theory, the paper proposes a regulatory system by integrating preventive legislation + dynamic supervision + technical traceability to emphasize the establishment of a tiered evaluation system for algorithmic impact and the introduction of third-party compliance audit mechanisms. By implementing an algorithm correction mechanism to avoid the risk of algorithmic discrimination, the paper innovatively proposes a "scenario-based attribution of responsibility for technology applications" model provides an institutional solution to balance technical efficiency and rights protection. The study also provides strong technical support and legal safeguards to address social issues and offers new perspectives and methods for scholars engaged in related research.

Keywords: Artificial intelligence; Anti-trafficking; Algorithm bias; Third-party audit; Technical application scenario attribution

1. Introduction

Human trafficking is a global governance problem that seriously violates the rights of citizens and threatens social security, regional distribution of global human trafficking cases as shown in Figure1. (a). The General Office of the State Council of China issued the "China Anti-Human

Trafficking Action Plan (2021-2030),” emphasizing the need to adhere to the work policy of "people-oriented, comprehensive management, prevention first, and combination of combat and prevention,” and vigorously supported the advancement of the anti-trafficking cause. Against this backdrop, the convergence of digitalization and artificial intelligence (AI) provides a new path for anti-trafficking. Through big data analysis, high-risk areas and potential victims can be identified in advance, which is in line with the concept of "prevention first.” However, technological empowerment also brings social risks such as data leakage and algorithmic bias.

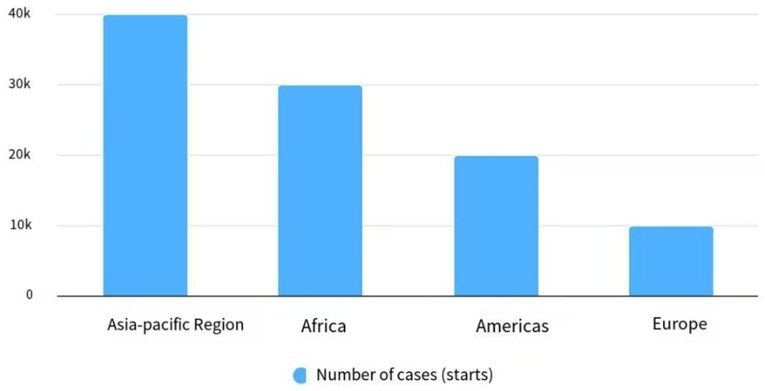

Figure1. (a) Regional distribution of global human trafficking cases (2022)

Source: By authors.

This study adopts empirical and normative research methods to focus on the artificial intelligence anti-trafficking system. On one hand, it reveals its technology-enabling logic in abduction-related intelligence research and judgment, evidence collection, and analysis. It also covers interrogation strategy generation through the bottleneck of manpower governance and “prevention-oriented, combining fight and prevention". On the other hand, it analyzes data abuse, algorithmic bias, fuzzy responsibility attribution, and other derivative risk issues, providing risk warnings and system construction.

The existing research focuses on the technical effectiveness of artificial intelligence anti-trafficking system, and the potential legal risks are not discussed enough, and the risk response strategies are not comprehensive and targeted. In terms of data security, the legal regulation of all aspects of the data life cycle is not perfect; in terms of algorithmic discrimination, the rights and interests of vulnerable groups are not fully considered; in terms of attribution of responsibility, there is lack of in-depth analysis of the complex relationship between multiple subjects. This study, however, focuses on the analysis of these risks, and put forward the construction of "preventive legislation dynamic supervision technology traceability" trinity of the regulatory system and other solutions to make up for the shortcomings of previous research.

This study aims to provide a feasible institutional design for the integration of artificial intelligence into the national anti-trafficking governance system, balance the relationship between technology and rights protection, and promote the transformation of anti-trafficking from "passive response" to "intelligent good governance". By adhering to the human-centered approach and aiming to protect the rights and interests of victims and maintain fairness and justice, this study provides theoretical and technical support for the anti-trafficking strategy, and revitalizes the value concepts of "benevolence" and "harmony".

2. Literature Review

In today's digital age, the clustering effect of big data, blockchain, and artificial intelligence has accelerated the construction of anti-trafficking systems, and its application and derivative risks have aroused the focus of academic attention and research. The following section will analyze the specific research of artificial intelligence application and risk from domestic and foreign aspects and profoundly grasp the dual variation of technical empowerment and legal regulation of artificial intelligence anti-trafficking system applications.

With the advent of the era of artificial intelligence, academic circles at home and abroad have set off a wave of exploration on the application of artificial intelligence and legal risk regulation. In early foreign countries, Briana. Browne (2017) discussed the challenges and responses of self-driving cars to the existing regulatory system in the United States, and put forward the needs of the new regulatory era.Drewsimshaw, Nicolas Terry, Krishauser, M.L. Cummings (2016) discussed the influence of medical robots on patient safety, privacy protection, medical-device supervision, and data protection laws. Nicholson Price II (2017) analyzed the application of artificial intelligence in the medical field and its legal issues, and emphasized the importance of safety performance and user privacy[1]. Gurney (2016) discussed the ethical and legal decision-making issues of self-driving vehicles and proposed six moral dilemmas and their legal and ethical considerations[2]. Chander (2017) discussed the problems of racial algorithms and proposed legal countermeasures for algorithm discrimination and algorithm bias[3].

With this development, theory is constantly evolving. LIU Yonghong, LI Wenying(2024) discussed the legal risks of generative artificial intelligence and its regulatory path, emphasized the importance of risk prevention and the principle of proportionality, and put forward a special legal framework and sandbox supervision measures. ZHENG Xi, ZHU Su-rong(2023) analyzed the legal risks and regulations of generative artificial intelligence, put forward risk prevention principles and ethical standards, and suggested breaking through the "vertical iteration" legislative model and strengthening the supervision and protection of data rights and interests. SONG Hua-jian(2024) studied the legal risks and governance path of generative artificial intelligence, and proposed a dynamic content feedback evaluation mechanism and data governance strategy to cope with the risks in language model training and content generation. WANG Y, PAN Y, YAN M, et al. (2023) discussed the content generation, challenges and solutions of ChatGPT and other generative artificial intelligence, and put forward the compliance and ethical issues of AI generated content.LI Z(2023) discussed the legal and ethical challenges of ChatGPT, especially the problems of random parrots and hallucinations, and put forward the legal and ethical framework of AI-generated content.

In the domestic field, Chen Siyu (2023) studied the legislative regulation mechanism of artificial intelligence algorithm infringement and proposed a path to realize the credible characteristics of artificial intelligence algorithms, such as interpretability, evaluability, and supervision. Wang Qinghua (2019) discussed the legal regulation path of artificial intelligence in a framework, emphasizing the applicability and reflection of the existing legal framework. Liu Yonghong and Li Wenying (2024) analyzed the legal risks of generative artificial intelligence and its regulation path, and put forward the importance of risk prevention and proportionality principle. Zheng Xi and Zhu Sunrong (2023) discussed the legal risks and regulations of generative artificial intelligence, and put forward the importance of compliance and ethical guidelines. Song Hualin (2023) studied the construction of artificial intelligence ethics from the perspective of rule of law, and put forward the legalization path of ethics. Sean (2024) explored the legal ethics form of artificial intelligence, and proposed a construction path of artificial intelligence ethics norms. Chai Ruijuan (2017) studied the overseas experience of supervision sandbox and its enlightenment, and put forward the innovative path of artificial intelligence supervision. Feng Jie (2022) discussed the legal path to improve the national security risk regulation of artificial intelligence, and put forward measures to reduce the immature and malicious application risk of artificial intelligence through law. Zeng Xiong, Liang Zheng and Zhang Hui (2024) analyzed the regulation path of artificial intelligence in EU and its enlightenment to China, and put forward the idea of risk classification and product regulation path.

From the above research, it can be seen that foreign scholars mainly focus on the application of artificial intelligence in specific fields and its legal issues, such as medical treatment and autonomous driving, while domestic scholars pay more attention to the legal regulation, ethical norms, and national security risks of artificial intelligence algorithms. Both of them deeply discussed the legal risks and regulatory paths of artificial intelligence technology from different angles, which provided an important theoretical basis and practical guidance for the application of artificial intelligence anti-trafficking systems. However, the current research has not been fully combined with the specific application scenarios of artificial intelligence anti-trafficking systems, and it is necessary to further explore its legal risks and regulatory paths in the field of anti-trafficking in the future to ensure the legal, safe, and effective application of technology.

3. Research Design

3.1 The Double Variant of Technological Empowerment and Legal Regulation

The field of abduction and trafficking crime governance is experiencing profound change, transforming from traditional human prevention and control to a paradigm empowered by intelligent technology[4]. AI technology, with its powerful data processing and analyzing capabilities, can quickly screen out key clues in the massive amount of data, providing strong support for case detection[5], which can accurately locate suspicious persons and potential clues, provide arrest solutions, and improve the abduction and trafficking case detection efficiency. As a result, the number of child trafficking cases in China has shown a downward trend year by year, as shown in Figure 2 (a).

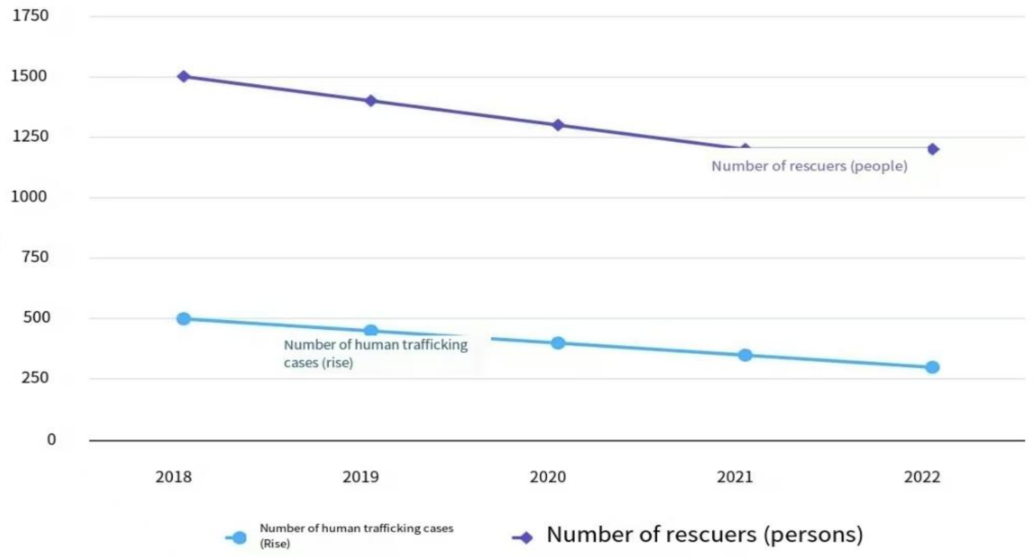

Figure 2. (a) Trend of Human Trafficking Cases in China (2018-2022)

While the application of technology has yielded results, we are also faced with a series of new problems, and this "Colling ridge’s dilemma" is particularly evident in the application of technology to the management of trafficking crimes[6]. At the initial stage of technological development and application, there is often a greater focus on maximizing the effectiveness of the technology and a relative lack of assessment and prevention of potential risks. When the technology is widely applied, once risk issues such as data leakage, algorithmic discrimination, and infringement of citizens' privacy arise, regulation and correction will not only require a huge investment in human, material, and financial resources, but may also face dilemmas such as difficulty in technological transformation and lagging legal regulations.

Although Article 14 of the current Anti-Trafficking Action Plan (2021-2030) mentions “strengthening scientific and technological anti-trafficking capacity building”, there is no specific specification for the application of technology. In practice, the AI anti-trafficking systems of public security organs in many places have been riddled with problems, with unclear boundaries for data collection and opaque algorithmic decision-making, making them prone to over-collecting personal information and infringing on citizens' privacy rights. For example, in the “face recognition miscarriage of justice” case in Hangzhou in 2023, the system mistakenly recognized normal parent-child interactions as abduction and trafficking, and the person concerned was inappropriately interrogated, not only suffering distress and mental harm, but also exposing the lack of rights protection in technological governance. Technology is intended to safeguard the rights and interests of citizens, but in the absence of effective regulation, it will harm the rights and interests of citizens.

3.2 Specific Applications of Artificial Intelligence Anti-Trafficking Systems in Human Trafficking

In the practice of combating human trafficking, public security organs across various regions have actively explored and innovated, forming a series of effective professional work models and mechanisms. Among them, the Zhangzhou 110 Missing Persons Studio has successfully established a new model for finding relatives that features rapid information transmission, precise investigation, efficient coordination, strong search capabilities, and extensive publicity. This model provides a robust support for addressing the urgent and difficult issues encountered during the process of finding relatives, effectively improving search efficiency and success rates. It serves as a vivid example of the practical application of artificial intelligence in anti-trafficking systems and paves the way for further development Exploring the deep integration and application of artificial intelligence technology in anti-trafficking efforts offers valuable insights. The following will delve into the specific applications of AI-based anti-trafficking systems from aspects such as intelligent analysis of trafficking-related intelligence, intelligent collection of evidence related to trafficking, and intelligent interrogation of suspects involved in trafficking.

3.2.1 Intelligent research and judgment of abduction-related intelligence

3.2.1.1 Constructing knowledge network models

Artificial intelligence system can collect and organize massive information, build a comprehensive and systematic trafficking crime database, extract a variety of child trafficking crime models and laws, build up a knowledge network model closely related to the core elements of "people, things, places, organizations", etc., and automatically analyze the new data with high efficiency and accurate comparison, providing powerful data support and technical guarantee for the work of preventing child trafficking. It provides strong data support and technical guarantee for the prevention of trafficking crimes[7] .

3.2.1.2 Facilitating retrospective investigations

In the face of abduction and trafficking cases that have already occurred, the AI anti-trafficking system shows its unique technical advantages, providing strong support for retrospective investigations[8] . The AI anti-trafficking system uses intelligent semantic recognition technology to analyze and label the data of solved cases and cases to be solved, and intelligently matches similar cases, providing valuable clues and clear investigation directions for investigators, and improving the efficiency and accuracy of investigation work[7].According to a report by the Gongjing District Procuratorate in Zigong City, Sichuan Province, the data from the district procuratorate showed that after the adoption of the "elementary interrogation system", the time for making records was reduced by 40 percent and the efficiency of evidence review was increased by 50 percent.

3.2.2 Intelligent collection of abduction-related evidence

3.2.2.1 Intelligent video system

Combining artificial intelligence with various types of surveillance equipment to build an intelligent video system has become an important development direction in the field of security. The system integrates various functions such as face recognition, iris recognition, voice recognition, age group recognition, gait recognition, expression recognition, object recognition, etc., and is able to quickly screen out the effective clips related to abduction-related cases in a massive video stream, which provides the detection and litigation of abduction-related cases with Strong support[9] .

3.2.2.2 Electronic data forensics

The intelligent system can generate case-specific investigative mind maps and business flow charts by summarizing the existing information and case-like data of the case, thereby providing clear forensic guidelines for investigators[10]. For network trafficking crimes, the intelligent system foregrounds electronic data forensic specifications into the intelligent forensic tools to ensure the legality and standardization of the forensic process. According to the official website of the Supreme

People's Procuratorate of the People's Republic of China, "Guangdong Qingyuan: Exploring the deep integration of artificial intelligence and procuratorial duties" was published on March 22,2025, the "Smart Prosecution" system of the procuratorate in Guangdong Qingyuan Province automatically compared contradictions between the suspect's confession and the victim's statement in a kidnapping case in 2024, and combined with electronic evidence (such as call records and transaction logs) to generate a visual analysis report, significantly reducing the investigation time; According to the Xiamen Law Information Network, the remote interrogation system (Fadu Shizhengtong) of the Xiamen Public Security Bureau, in a cross-border kidnapping case in 2023, used intelligent transcript analysis to automatically link the suspect's past criminal records, revealing similar modus operandi to multiple unsolved kidnapping cases, ultimately leading to the solving of related cases. Through a deep analysis of case elements, these intelligent systems construct a complete picture, which provides accurate guidance for investigators.

3.2.3 Intelligent interrogation of abduction suspects

In the investigation of abduction-related cases, the intelligent interrogation system provides all-round and intelligent auxiliary support for interrogators by integrating a variety of advanced technologies, which significantly improves the efficiency and quality of interrogation work[11]. The intelligent interrogation system identifies suspects through identity information comparison and other technologies, and uses big data to analyze behavioral habits and character traits to provide reference bases for interrogators. According to the official website of Jiangxi Public Security Department, "Wuxi Public Security aims at practical combat and plans reform to cultivate new quality combat effectiveness with integrated attack and defense mechanism" in 2023, when investigating a kidnapping case in Wuxi, Jiangsu Province, police used a smart interrogation system to compare the mobile phone signal trajectory of the suspect with the surveillance picture through big data, and successfully locked and arrested the criminal suspect in just 24 hours, which improved the efficiency by nearly 60% compared with traditional investigation methods.

During interrogation, the system monitors the suspect's emotional and psychological dynamics in real time and helps interrogators adjust their strategies by capturing subtle changes. The "Elementary Examination System" used by the Gongjing District Procuratorate of Zigong City, Sichuan Province, can automatically identify psychological indicators such as anxiety and lying tendencies of suspects. In the kidnapping and extortion case of 2024, the system helped interrogators break through the suspect's psychological defenses within three hours through real-time psychological monitoring, leading to a complete confession of the criminal facts.

After the interrogation, the system identifies key elements in the transcript through semantic extraction function and builds a case analysis model to compare and analyze the key information with the data in the case cloud system, to dig out potential clues and evidence. According to an article published on the official website of Ningbo Municipal Justice Bureau, when solving a cross-border kidnapping case, the police in Ningbo used the "psychological profiling" function of the intelligent interrogation system to analyze the suspect's social network behavior patterns, accurately predict their hiding place, and ultimately successfully rescue the victim. Post-assessment showed that the system helped the police complete the investigation of cross-border cases within 72 hours, significantly reducing the average time required for similar cases.

3.3 Spectrum of Legal Risks in the Application of Artificial Intelligence Anti-Trafficking

Systems

3.3.1 Data security risks

The operation of the artificial intelligence anti-trafficking system relies on a large amount of data, and there are many security risks in the collection, storage, transmission, and use of this data[9]. On one hand, data may be illegally stolen or leaked[12], leading to the exposure of personal privacy and causing secondary harm to victims and their family members. However, the data may be over collected and misused.

3.3.1.1 Crisis of legitimacy of data collection

In the existing AI anti-trafficking systems, the legality of data collection has become increasingly prominent. From the perspective of public trust, excessive collection of sensitive information without the express consent of guardians may lead to a decline in public trust in the anti-trafficking system. From the perspective of technology applications, anti-trafficking systems should strictly comply with laws and regulations in the process of data collection to ensure the legality and compliance of data collection.

3.3.1.2 Regulatory blind spots in data sharing

In the current digital era, data sharing has become an important means of promoting social governance and innovation in public services. However, there are still some regulatory blind spots in the data-sharing process, and the lack of a legal basis for authorization is particularly prominent. The legal framework for cross-sectoral data sharing is not yet perfect, and there is a lack of unified regulatory bodies and standards, which makes it difficult to effectively monitor the entire data-sharing process and increases the risk of data leakage and illegal use.

In addition, data sharing has data security issues of leakage, tampering, and misuse, which will lead to the exposure of sensitive information when data are not effectively encrypted and desensitized.

3.3.1.3 Security risks of data storage and transmission

In the era of digitization, data have become a key resource, but the security risks during storage and transmission cannot be ignored. Data storage systems may be subject to crises such as hacker attacks, leading to data leakage or tampering[13]. These security risks not only threaten personal privacy but also affect the normal operation of the anti-trafficking system. There is an urgent need to strengthen the security measures for data storage and transmission.

3.3.1.4 Risk of misuse of data use

Artificial Intelligence has over-collection and misuse risks for data collection and processing[9]. The problem of over-collection and misuse of data has become increasingly prominent in today's era, and the improper use of data is not only limited to privacy leakage but may also give rise to a series of other risks; if the collected personal information is used for commercial purposes unrelated to anti-trafficking or other illegal purposes, it will seriously violate the legitimate rights and interests of individuals[14].

3.3.2 Questioning the legitimacy of algorithmic decision-making

The legitimacy of algorithmic decision making in the application of artificial intelligence anti-trafficking systems has been questioned in many ways. These questions not only affect the fairness and transparency of the algorithm but may also lead to misjudgment or neglect of certain groups. To ensure the legitimacy of algorithmic decision-making, it is necessary to strengthen the standardization of data management and algorithm design to ensure the fairness, transparency, and interpretability of algorithmic decision-making.

3.3.2.1 Discrimination reinforcement by characteristic labeling

In the process of rapid development and wide application of artificial intelligence technology, the problem of discrimination reinforcement of feature labels has gradually come to the fore[14], and has become an important factor restricting the fairness and impartiality of algorithms. According to the National Institute of Standards and Technology (NIST), the training data for AI algorithms have significant geographical and group biases [15], as shown inFigure3. Algorithmic discrimination comes mainly from two aspects. First, there is a bias in the training data. Algorithms are trained or run-on real-world datasets, and real-world invisible biases are very easy for algorithms to capture and learn from [16]. The other flip side is the input of the human subjective bias in the algorithmic decision-making process. At every step, from data collection to algorithmic decision making, there can be inputs of human subjective bias.

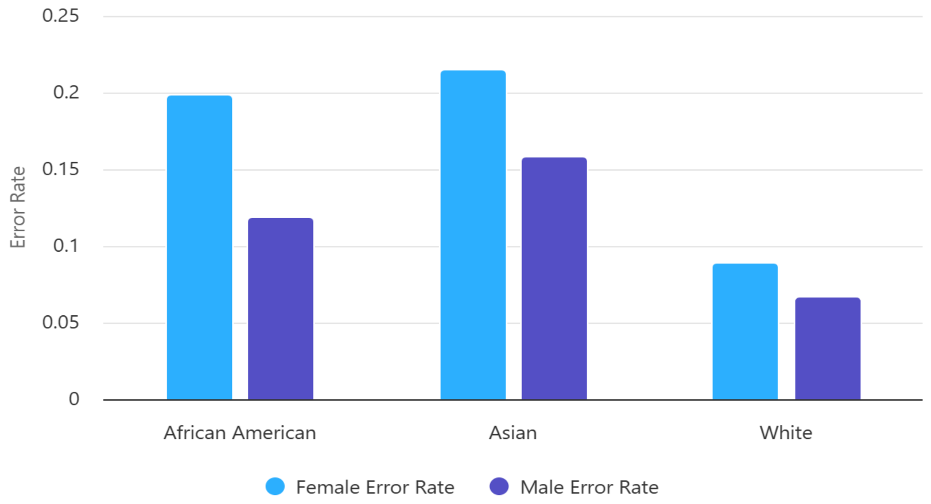

Figure 3. Error Match Rates by Race and Gender

3.3.2.2 Insufficient transparency and interpretability of decision-making

In the application of AI anti-trafficking systems, the "algorithmic black box" problem arising from the complexity and opacity of algorithms poses a significant challenge to the legitimacy of the administrative process[17]. Algorithmic decision-making systems often lack audit trails designed to document the entire automated decision-making process. This shortcoming not only undermines the right to information of administrative counterparts, making it difficult for them to understand the basis and process of decision-making, but may also lead to mistrust of the results of algorithmic decision-making, which in turn affects their acceptance of the overall administrative process and cooperation with.

3.3.2.3 Balancing fairness and efficiency

In the application of AI anti-trafficking systems, the justification of algorithmic decision-making faces the problem of balancing fairness and efficiency[18]. The transient nature of algorithmic decision making is an inherent limitation. Algorithmic transience makes it difficult for administrative subjects to perform human intervention and correction in the decision-making process, which may lead to more serious bias in the results of algorithmic decision-making. However, algorithmic decision-making may also lead to fairness issues. Algorithmic efficiency gains may occur at the expense of fairness. For example, algorithmic decision making may over monitor certain groups while ignoring other areas that are equally at risk.

3.3.3 Liability attribution risk

The risk of responsibility attribution in the artificial intelligence anti-trafficking system is mainly reflected in the difficulty of composite responsibility subjects, abuse of technology-neutral defenses, and aggravation of responsibility-shifting situations[19]. To ensure the effective operation of and public trust in the AI anti-trafficking system, it is necessary to clarify the responsibility of each subject, establish a sound accountability mechanism, and ensure that victims can obtain effective relief.

3.3.3.1 Composite responsible body difficulties

The operation of an AI anti-trafficking system involves multiple subjects, including algorithm developers, data suppliers, and system operators, which play different roles and assume different responsibilities in the system. The boundaries of responsibilities among different subjects are blurred, with algorithm developers responsible for designing and writing algorithms, data suppliers providing data, and system operators responsible for the daily operation and maintenance of the system. When the system fails, it is difficult to determine whether it is a problem of algorithm design, data quality, or system maintenance; the allocation of responsibilities among subjects lacks a clear legal basis. Existing laws and regulations do not make specific provisions on the attribution of responsibility for AI systems, resulting in the allocation of responsibility often being bogged down in practice, which not only increases the complexity of litigation, but also makes it more difficult for victims to obtain compensation

3.3.3.2 Abuse of the defense of technological neutrality

In the case of artificial intelligence anti-trafficking systems, technology providers often invoke Article 41 of the Product Quality Law to claim exemption from liability, leading to a "technical black box" in the determination of liability. The abuse of this defense strategy allows technology providers to avoid liability easily, thereby affecting the likelihood of victims receiving compensation. If technology providers can easily avoid liability, it will be difficult for victims to obtain due compensation, which not only jeopardizes the legitimate rights and interests of victims but also weakens the authority of the law.

3.3.3.3 Increased shifting of responsibilities

In an AI anti-trafficking system, there can be a shift in responsibility among various subjects. The algorithm designer may shift the responsibility to the data provider on the grounds that the data are incorrect, while the data provider may shift the responsibility to the manager of the AI on the grounds that there are problems in the daily management and maintenance of the AI, and the manager and maintainer may also point the finger at the algorithm designer, claiming that it is the faulty design of the algorithm, which has led to the final wrong result. Such a shift in responsibility increases the complexity of attribution and affects the likelihood that victims will be able to obtain redress.

4. Results and Discussion: Path to Regulate the Legal Risks of the Application of Artificial Intelligence Anti-Trafficking System

Given the complexity and real-world harm of legal risks in the application of AI antitrafficking systems, it is imperative to establish comprehensive and targeted regulatory approaches to address these challenges. The following sections will elaborate on specific regulatory strategies and measures from the perspectives of building new systems, establishing new mechanisms, and constructing new models, providing institutional safeguards for the lawful and efficient use of AI anti-trafficking systems.

4.1 Constructing a Trinity of the Regulatory System of "Preventive Legislation + Dynamic Supervision + Technical Traceability.”

4.1.1 Preventive legislation

To clarify the applicable legal norms of AI anti-trafficking systems in the legislation and to make detailed regulations on the collection, storage, use, and sharing of data to ensure the safety and lawful utilization of data[20]. To ensure the healthy and orderly development of AI anti-trafficking systems, we need to do a good job of risk prevention in the legal framework, adhere to the principle of proportionality, and do a good job of sandbox regulatory measures[21], break through the traditional "vertical iteration" legislative model, and strengthen the protection of data rights and interests[22]. The establishment of "GB/T35273C class" biometrics special standards, minors data protection level to the highest level, is an important measure to protect data security. In the process of research and development, application, and management of AI anti-trafficking systems, technical ethics should not be ignored. Technical ethical guidelines should be formulated, the basic principles to be followed in the application of technology should be clarified, and a sound review mechanism for scientific and technological ethics should be established to ensure that the application of AI anti-trafficking systems meets social ethical standards[16].

4.1.2 Dynamic regulation

During the development, use, and management of the artificial intelligence anti-trafficking system, dynamic supervision is an important means of ensuring the safe and effective operation of the system. It is necessary to speed up the establishment of specialized regulatory bodies with independent law enforcement powers and professional technical capabilities to conduct regular inspections and assessments of the system’s operation and to be responsible for supervising and managing the full life cycle of the system. A scientific and reasonable evaluation mechanism should be established to regularly assess the performance, effectiveness, and safety of the system, the results of which should be used as the basis for system improvement and optimization and should also be made public and subject to public supervision. Focusing on risk monitoring and early warning during the operation of the system, real-time monitoring of the system's operational status should be carried out through the use of modern information technology, so that potential risks and problems can be detected and dealt with in a timely manner.

4.1.3 Building a mechanism for tracing technology applications

In the application process of the artificial intelligence anti-trafficking system, the construction of the technology application traceability mechanism is an important means to ensure that the system is transparent and trustworthy[23]. The use of blockchain technology for the entire process of data collection, algorithmic decision-making, and the resultant application is the key to realizing an auditable aftermath. Building a complete operational log system is an important part of the traceability mechanism of technology applications. The algorithmic decision-making traceability requires the algorithm of the AI anti-trafficking system to have a certain degree of transparency, and records each decision-making process in the operation process, including the original data input, intermediate calculation results, and final decision output. System responsibility traceability is an important part of the technology application traceability mechanism, and should clarify the responsibility of system developers, data providers, users, regulators, and other participants.

4.2 Establishment of A Hierarchical System for Assessing the Impact of Algorithms

To effectively manage the risk of algorithms in the AI anti-trafficking system and ensure fairness, transparency, and reliability of the algorithms, an algorithm impact grading assessment system[24] can be established.. The system grades and assesses algorithms according to their application scenarios, influence scope, and potential risks and implements differentiated regulatory measures.

The classification of algorithm levels should be led by the Ministry of Public Security, in conjunction with the Supreme People's Procuratorate and Supreme People's Court, based on risk impact levels and technical complexity standards. This process should revolve around the principles of compliance, fairness, transparency, dynamism, and user protection, ensuring consistency and authority at both legal and enforcement levels. Furthermore, when assigning algorithm classification levels, it is essential to involve technical experts, legal experts, and non-governmental organizations, fully considering multiple dimensions, such as technology, law, and society, to make the classification standards more scientifically sound. Simultaneously, administrative supervision mechanisms should be introduced, and technical audit institutions should establish an AI algorithm supervision committee to oversee the overall planning and guidance of AI algorithm supervision. On this basis, the primary responsibilities of algorithm developers should be clarified and a user joint liability system should be established. A dedicated complaint and remedy channel should also be set up to provide effective avenues for individuals or organizations whose rights have been infringed due to improper algorithm behavior, fully protecting their legitimate interests.

4.2.1 Management of low-risk algorithms

Low-risk algorithms refer to those in the artificial intelligence anti-trafficking system, mainly dealing with open, low-sensitivity data that do not directly affect personal rights and interests, high error tolerance, and relatively simple data processing logic. For example, face recognition algorithms have less impact on individual rights, interests, and public social interests. The management measures for such algorithms can be managed by the filing system, requiring developers to keep complete algorithm logs and operation records, such as detailed records of the algorithm's usage time, usage scenarios, data sources, and other information to ensure that they operate within the compliance framework, thus protecting the legitimate rights and interests of users and the public.

4.2.2 Management of medium risk algorithms

A medium-risk algorithm refers to an algorithm that involves a certain degree of personal privacy data processing in an artificial intelligence anti-trafficking system, such as personally identifiable data, and the data processing logic is relatively complex. It may have a certain impact on individual rights and interests as well as social and public interests, and may cause compensable losses. For example, predictive algorithms involving individual behavior analysis, which need to collect and analyze individual behavior data over a certain period, have a relatively large impact on individual rights and interests and the public interest of society. Therefore, in addition to record management, it is necessary to conduct safety audits and technical tests by professional safety audit institutions on a regular basis (quarterly or semi-annual), introduce independent third-party technical assessments, and establish error correction mechanisms to ensure that they are under control. Run.

4.2.3 Management of high-risk algorithms

High-risk algorithms, in the artificial intelligence anti-trafficking system, refer to those algorithms that deal with highly sensitive personal data (such as personal biological characteristics, financial information, health information, etc.), which may involve complex operations such as deep mining and analysis. Once the algorithm is wrong or abused, it directly affects personal freedom and other major rights and interests, and the error may cause irreversible and serious consequences, mainly based on the sensitivity of the data, complexity of data processing, and potential harm to individuals and society. For example, data-mining algorithms. Such algorithms can be regulated and constrained by management measures such as establishing a pre-administrative approval system, running a monitoring system in real time, conducting annual comprehensive security assessments, and strictly limiting application scenarios.

The main bodies responsible for the implementation of high-risk algorithm supervision are the relevant government departments (such as data supervision departments and artificial intelligence supervision departments). Strict approval and regulatory measures should be implemented for this algorithm. Before the algorithm is put into use, a professional evaluation team organized by the relevant government departments must conduct a comprehensive evaluation, which can be put into use after passing. At the same time, it limits its application scenarios and clarifies the scope of its application to avoid risks arising from improper use.

4.3 Introduction of A Third-Party Compliance Audit Mechanism

In today's digital era, the introduction of a third-party organization with specialized technical capabilities and legal knowledge can conduct comprehensive and objective assessments and audits of the system's data security, algorithmic fairness, and attribution of responsibility to ensure sound operation and legal compliance.

During the audit process, third-party organizations need to focus on the legality of the data source to ensure that the acquisition and use of data comply with the requirements of laws and regulations and to avoid legal risks and ethical disputes arising from improper data sources. Technical ethical indicators, such as whether the algorithms are biased, are also important elements of the audit. In addition, the openness and transparency of the audit results are crucial. Third-party organizations should disclose the audit results to society and accept public supervision. A third-party agency can also draw on the existing third-party monitoring and assessment mechanism for compliance of enterprises involved in the audit process to ensure the standardization and effectiveness of the audit work.

The EU's "Artificial Intelligence Act" requires that the development and use of high-risk artificial intelligence systems must be audited by an independent third-party agency to ensure that they comply with EU safety and ethical standards. This mandatory audit mechanism provides a strong guarantee for the reliability and security of artificial intelligence systems. The Algorithms Accountability Act focuses on the fairness and transparency of algorithms, requiring companies to conduct internal and external assessments of automated decision-making systems to identify and correct possible biases and discrimination. This evaluation mechanism focuses on the input and output processes of the algorithm to ensure fairness in the algorithm's decision-making.

Compared to the above-mentioned international experience, the introduction of a third-party compliance audit mechanism in China has certain similarities and can be used as a reference. All supervise and evaluate artificial intelligence systems through third-party forces independent of development and operation entities to ensure their legal and compliant operations. However, certain differences were also observed. The European Union pays more attention to the overall security and reliability of the system, while the United States focuses on the fairness and transparency of the algorithm. The introduction of a third-party compliance audit mechanism in my country can be adjusted and optimized based on international experience according to the country's national conditions and the specific characteristics of the artificial intelligence anti-trafficking system. For example, according to China's legal system and regulatory needs, we can clarify the qualifications and responsibilities of third-party audit institutions, formulate audit standards and procedures suitable for China, and focus on the dual review of algorithm fairness and data security.

4.4 Innovative "Scenario-Based Attribution of Responsibility for Technology Applications" Model

In order to effectively deal with the problem of responsibility attribution in the application of AI anti-trafficking system, we have innovatively put forward the model of "Scenario-based Attribution of Responsibility for Technology Application," which reasonably determines the attribution of responsibility based on the specific scenarios of the application of the technology, the behavioral performance of each party, the degree of fault and other key factors, so as to provide the victims with a clear path to relief, and prompt each party to be more cautious and responsible in the application of the technology. This will provide a clear remedy for victims and encourage all parties to be more cautious and responsible for the process of technology application[25].

In practical applications, the failure of a face recognition system to recognize a missing person, for example, should be considered in a comprehensive manner using a variety of factors. The technical level of the system developer, accuracy of the algorithm, and adequacy of data training are directly related to the recognition capability of the system. The maintenance and management of the operator include regular updating of the system, troubleshooting, and emergency response to abnormal situations. Users' operational standardization, whether they operate according to the prescribed procedures, whether they pay enough attention to system prompts, etc., will affect the actual effect of the system. The data quality of data providers, such as data accuracy, completeness, and timeliness, is directly related to the recognition effect of the system. By comprehensively considering these factors and assuming the corresponding legal responsibility according to the degree of fault and the proportion of contribution to the result of the damage, this attribution of responsibility not only embodies fairness and reasonableness but also effectively solves the problem of attribution of responsibility and avoids the phenomenon of shifting responsibility.

4.5 Improve the Algorithm Deviation Correction Mechanism

4.5.1 Fairness-aware algorithm design

The fairness constraint algorithm is introduced into the anti-trafficking intelligent system to ensure the fairness of the model's decision making among different groups (e.g., different skin colors, ages, and regions) by optimizing the target function and adjusting the feature weights[26]. For example, using an adversarial training mechanism, a fairness loss function is added to the algorithm to force the model to learn feature representations that do not rely on sensitive attributes (e.g., race and gender). Additionally, the model was dynamically monitored in conjunction with a fairness assessment tool to identify and correct potential biases.

4.5.2. Diversified training data construction

Through multisource data fusion, synthetic data generation, and other means, a training dataset covering different groups was constructed[27]. For example, multidimensional data such as the public security department's missing persons database, social media images, and cross-age face databases, and using data enhancement techniques (e.g., oversampling and interpolation) to balance minority samples. Simultaneously, a data audit mechanism is established to regularly check the balance of data distribution and avoid sample selection bias.

4.5.3. Ethical review and interdisciplinary collaboration

Establish an ethics review committee composed of computer scientists, ethicists, and legal experts to evaluate compliance with the entire process of AI system design, training, and deployment[28]. Simultaneously, we should promote interdisciplinary cooperation and integrate research results in sociology, psychology, and other fields into algorithm design to ensure that the application of technology conforms to social ethical norms. For example, the introduction of "fairness impact assessment" in the system development phase predicts the potential harm that the algorithm may cause to a specific group.

4.5.4. Synthetic data and edge computing optimization

In view of the problem of data scarcity, high-quality training data are synthesized by generating adversarial networks and other technologies to fill the data gaps for specific groups[29] (such, ethnic minorities and special regions). Simultaneously, combined with edge computing technology, some model reasoning tasks are sunk to terminal devices, reducing dependence on centralized data and reducing the risk of data leakage.

5. Conclusion

The application of artificial intelligence technology in the field of anti-trafficking has important practical significance and broad development prospects but also faces many legal risks. The focus of this paper is to deal with new types of legal risks such as data security, algorithmic discrimination, and attribution of responsibility derived from the application of AI, and effectively regulate the application of AI technology through the construction of a three-in-one regulatory system of "preventive legislation + dynamic supervision + technological traceability", the establishment of an algorithmic impact grading and assessment system, the introduction of a third-party compliance auditing mechanism, and the innovation of a "technological application scenario-based attribution of responsibility" model. By constructing a three-in-one regulatory system of "preventive legislation, dynamic supervision and technical traceability," establishing a hierarchical assessment system for the impact of algorithms, introducing a third-party compliance audit mechanism, innovating the model of "attribution of responsibility for technology application scenarios" and algorithm correction mechanism, we can effectively regulate the legal risks of the application of AI anti-trafficking systems, balance the relationship between technological efficiency and the protection of rights, and promote the healthy and orderly development of AI technology in the field of anti-trafficking, so that we can provide strong technological support and legal protection for the resolution of the social problem of human trafficking.

From the perspective of future prospects, with the continuous development of artificial intelligence technology, its application in the field of anti-trafficking will be more extensive and in-depth. By continuously optimizing the above-mentioned regulatory measures and models, it can promote the healthy development of artificial intelligence technology in the field of anti-trafficking and provide strong technical support and legal protection for solving social problems. In future development, we should pay close attention to the development and application practice of artificial intelligence technology, constantly improve relevant laws, regulations, and regulatory mechanisms, and timely adjust and improve regulatory strategies according to the new characteristics and problems of technological development, so as to ensure that artificial intelligence technology always serves the well-being of human society and plays a greater positive role in social public welfare undertakings such as anti-trafficking, helping to build a safer, harmonious, and just social environment. In future development, we should pay close attention to the development dynamics and application practice of AI technology, and constantly improve the relevant laws, regulations, and regulatory mechanisms to ensure that AI technology always serves the well-being of human society.

Acknowledgements

This study was supported by the proposed project "Personality Assessment Mechanism for Special Education of Juvenile Offenders" (Project NO. JAS24052) under the 2024 Fujian Province Education System Philosophy and Social Science Research Program. I would like to extend my special thanks to Professors Hong Liou and Ye Xiuwen for their heartfelt guidance, assistance, and answers during the writing process, which helped resolve many shortcomings in the paper.

The heart is gently touched, and the water is clear with the sand bright. Here comes the end of the writing, and the pen is put down to conclude.

References

[1] Price, P.W. Artificial intelligence in healthcare: Promises and challenges. Richmond Journal of Law & Techn ology, 2017, 23(2), 1–35.

[2] Gurney, J.K. The ethics of driverless cars. ACM SIGCAS Computers and Society, 2016, 45(3), 179–184. DO I: 10.1145/2874239.2874265.

[3] Chander, A. The racist algorithm. Michigan Law Review, 2017, 115(6), 1023–1045.

[4] Wu, H. Research on countermeasures for the governance of network crimes of trafficking in women and chil dren. Network Security Technology and Application, 2023, (11), 142–144.

[5] Shen, X., Liu, H. and Wang, H. Application of generative artificial intelligence technology in extracting and summarizing key information of scientific and technical journal papers. China Science and Technology Journ al Research, 2025, 36(01), 37–43.

[6] Tong, Y. Out of the Colingridge dilemma: Dynamic regulation of generative artificial intelligence technology. Journal of Shanghai Jiao Tong University (Philosophy and Social Science Edition), 2024, 32(08), 53–67. D OI: 10.13806/j.cnki.issn1008-7095.2024.08.004.

[7] Zhang, L. and Yang, Z. Research on the risk prevention of criminal intelligence research and sentencing erro rs. Crime Research, 2024, (04), 41–50.

[8] Lee, J. Access to finance for artificial intelligence regulation in the financial services industry. European Busi ness Organization Law Review, 2020, 21, 731–757.

[9] Sun, Q. Legal risks and their prevention in the application of artificial intelligence anti-bullying system. Teaching and Management, 2025, (05), 11–14 + 19.

[10] Chen, Z. Legal reasoning logic of intelligent adjudication system. Journal of Sichuan Normal University (Social Science Edition), 2024, 51(02), 68–76 + 201. DOI: 10.13734/j.cnki.1000-5315.2024.0304.

[11] Wu, Y. and Liu, W. Status quo, dilemma and breakthrough in the application of artificial intelligence in investigative interrogation. Journal of Guangxi Police Academy, 2023, 36(01), 50–59. DOI: 10.19736/j.cnki.gxjcxyxb.2023.0106.

[12] Chen, B. Science and technology ethics and rule of law construction of artificial intelligence application. People’s Forum, 2024, (12), 66–70.

[13] Wang, Y., Pan, Y. and Yan, M. [et al.] A survey on ChatGPT: AI generated contents, challenges, and solutions. IEEE Open Journal of the Computer Society, 2023, 4(1), 280–302.

[14] Song, H. Legal risks and governance path of generative artificial intelligence. Journal of Beijing Institute of Technology (Social Sciences Edition), 2024, 26(3), 134–143.

[15] Huang, B. Risk types and legal regulation of training data of artificial intelligence large model. Politics and Law, 2025, (01), 23–37.

[16] Wang, Y., Pan, Y., Yan, M., Su, Z. and Luan, T.H. A survey on ChatGPT: AI generated contents, challenges, and solutions. IEEE Open Journal of the Computer Society, 2023, 4(1), 280–302. DOI: 10.1109/OJCS.2023.3300321.

[17] Wang, D. Limitations of algorithmic transparency mechanisms and their overcoming. Journal of East China University of Politics and Law, 2025, 28(01), 33–44.

[18] Feng, X. and Li, D. The challenge of algorithmic decision making to administrative due process and its response. Academic Exchange, 2024, (05), 59–72.

[19] Zhang, L. The dilemma of liability attribution and solution path of judicial application of artificial intelligence. Contemporary Law, 2023, 37(05), 100–111.

[20] Zhang, Y. On the synergistic governance of artificial intelligence policy and law. Oriental Law, 2024, (5), 187–200. DOI: 10.3969/j.issn.1007-1466.2024.05.014.

[21] Liu, Y. and Li, W. Legal risks of generative artificial intelligence and their regulations. Journal of China West Normal University, 2024, 37(9), 0–0.

[22] Zheng, X. and Zhu, S. Legal risks and regulations of generative artificial intelligence. Changbai Journal, 2023, 36(6), 80–88.

[23] Tian, Y. Dilemma and improvement path of blockchain technology in electronic data deposit. Henan Science and Technology, 2022, 41(09), 151–154. DOI: 10.19968/j.cnki.hnkj.1003-5168.2022.09.034.

[24] Zhang, X. The construction mechanism of algorithmic impact assessment system and China's program. Legal Business Research, 2021, 38(02), 102–115. DOI: 10.16390/j.cnki.issn1672-0393.2021.02.008.

[25] Liang, Y.-G. Scenario based classification and attribution of tort liability of generative artificial intelligence service providers. Journal of Shenzhen University (Humanities and Social Sciences Edition), 2024, 41(05), 115–124.

[26] Wang, M., Deng, W. and Su, S. Research progress of fairness in image recognition. Chinese Journal of Image and Graphics, 2023, 28(03), 612–624.

[27] Wang, Y., Wang, D. and Wang, H., et al. Algorithmic equity: Logic and governance of algorithmic bias in educational artificial intelligence. Open Education Research, 2023, 29(05), 37–46. DOI: 10.13966/j.cnki.kfjyyj.2023.05.004.

[28] Hanna, M., Pantanowitz, L., Jackson, B., et al. Ethical and bias considerations in artificial intelligence (AI)/

machine learning. Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc, 2024, 38(3), 100686.

[29] Yinlong, J., Chen, X. and Zhongrui, J., et al Data enhancement scheme in federated learning based on conditional generative adversarial networks. Computer Application, [online], 2025 04 20, 1–16. http://kns.cnki.net/kcms/detail/51.1307.TP.20250314.1702.012.html.