Prediction-Early Warning-Prevention: The Value Orientation, Future Dimension, and Practical Approach of Artificial Intelligence-Enabled Safety Management

Zhiqiang Wen, Chengyue Yang*

Tianjin Normal University, Tianjin, 300387, China
**Corresponding author: 18056108525@163.com*

Abstract: *Against the backdrop of digital transformation and increasingly complex risks, artificial intelligence enables safety management to shift from passive response to active prevention. Artificial intelligence employs technologies such as machine learning, edge computing, and digital twins to enhance the accuracy of risk prediction and the sensitivity of early warning response, thereby constructing a closed loop of the prevention system. In terms of technological application, challenges such as algorithmic bias and human-machine trust crises emerge. Therefore, this study proposes a "three-dimensional collaboration" framework that includes formulating ethical guidelines for artificial intelligence, ensuring human control over key decisions, and promoting public safety literacy education. In the future, attention should be directed toward technological integration, scenario expansion, and governance innovation, driving the intelligent and systematic upgrading of safety management.*

Keywords: *Artificial Intelligence; Safety Management; Prediction-Early Warning-Prevention*

1. Introduction

As society places increasing emphasis on safety management, empowering safety management with artificial intelligence has become an important development direction. With the rapid development of a new round of technological revolution and industrial transformation, emerging technologies such as artificial intelligence are flourishing, which greatly enhances human ability to understand and transform the world[1]. In the context of the interweaving of accelerating digital transformation and increasingly complex risks, global safety management is undergoing a historical shift from passive response to active defense. Traditional safety management models rely on manual inspection, experience-based judgment, and static rules, making it difficult to cope with unpredictable risks such as cyberattacks, emergencies, and natural disasters. The intervention of artificial intelligence technology has achieved three major breakthroughs:

a: Improvement of prediction capability. By using machine learning algorithms to deeply mine data, artificial intelligence can recognize the evolution patterns of risks. A community smart health service platform, relying on advanced big data technology and intelligent analysis models, provides personalized and convenient health management services to community residents, and the platform's accuracy in predicting health risks reaches 87%[2]. This breakthrough enhances the level of protection in terms of life safety.

b: Faster early warning. The combination of computing and data analysis enables an exponential increase in the response speed of early warning. The early warning system of an engineering project can fully consider the complexity and uncertainty of the project, and by building a multi-dimensional, multi-level early warning indicator system, it achieves comprehensive monitoring of project risks[3]. This breakthrough reduces economic losses.

c: Formation of a closed-loop prevention system. The combination of digital twin technology and reinforcement learning drives preventive measures to transition from experience-driven to data-driven. Taking this community smart health service platform as an example, the platform is designed based on a modular architecture, with each module connected to one another, forming a closed-loop health service system. From data collection, sorting, analysis to application governance, it builds a comprehensive health service support system[2]. This breakthrough improves the efficiency of risk event handling.

In the future, artificial intelligence-enabled safety management still needs to overcome three major technical bottlenecks:

a: The construction of a multi-module perception network. It integrates data sources such as satellite remote sensing, drone inspection, and social media public opinion, and develops a multi-module association analysis engine.

b: A breakthrough in autonomous evolutionary decision-making systems. It allows artificial intelligence models to repeatedly simulate exercises like receiving a vaccine. Just as a virus and antivirus software compete with each other, this training enables the defense system to become smarter with each battle and automatically learn to cope with various new types of attacks.

c: Innovation in human-machine collaboration mechanisms. The operation interface uses conspicuous buttons or voice prompts to ensure that key decisions can only be executed after a human presses the confirmation button, just as reliable as when an elevator overload requires passengers to press the button to close the door.

However, behind technological empowerment lie deep-seated contradictions, such as algorithmic bias and discrimination, which create certain unfair dilemmas. A trust crisis exists in human-machine collaboration, as people lack sufficient confidence in the system's processing capabilities. Over-reliance on the system may lead to an inability to think independently and handle problems. Based on this, this study proposes a "three-dimensional collaboration" implementation framework:

a: It formulates ethical guidelines for artificial intelligence safety applications, establishes an algorithm filing and impact assessment system, and strives for fairer handling of related issues.

b: It ensures that human experts retain ultimate control and the right of interpretation over key decisions, eliminating the possibility of system arbitrariness.

c: It implements a nationwide "safety literacy" education program, integrating artificial intelligence risk awareness into the national education system to prevent excessive reliance.

The technological revolution centered on artificial intelligence, through data collection, algorithm execution, and intelligent management, is reshaping the value logic and practical path of safety management. From improving the accuracy of earthquake early warning systems to real-time monitoring of hidden dangers in major internet security, and from risk simulation in smart cities to intelligent decision-making in emergency rescue, artificial intelligence is building a safety protection network covering the entire spectrum of "prediction-early warning-prevention". Empowering safety management with artificial intelligence is not only a critical revolutionary improvement in technical efficiency but also a cognitive revolution in humanity's response to uncertain risks. In the dynamic balance of efficiency and ethics, as well as innovation and safety, it is necessary to create an intelligent safety ecosystem featuring "precision in prediction, agility in early warning, and systematization in prevention". Future research should focus on cutting-edge fields such as human-machine hybrid intelligence and multi-module learning, injecting more sustainable intelligent momentum into safety management.

2. Value Orientation

On the basis of a literature review, this paper reveals the value progression of artificial intelligence in safety management from passive response to active intervention. In the future, artificial intelligence will focus on areas such as multi-modal data fusion and autonomous decision optimization, thereby promoting the formation of a practical paradigm of human-machine collaborative governance across the entire chain of "monitoring-response-governance". The deep application of artificial intelligence technology is transforming the underlying logical architecture of safety management, and its core significance lies in driving safety management from "experience-driven" to "data-driven". Existing research has identified three core value orientations:

2.1 Precision of Risk Prediction

Artificial intelligence uses multi-modal data fusion to greatly enhance the accuracy of risk prediction. The development and utilization of public data resources should combine market-driven forces with government regulation, using policies to provide guidance. By adding the data industry to the industrial structure adjustment guidance catalog, it drives technological innovation, cultivates high-level enterprises engaged in data element businesses, and supports the development of social organizations and industry alliances such as data industry associations and societies[4]. Furthermore, artificial intelligence-related technologies such as natural language processing, image recognition, and semantic analysis mainly focus on identifying structured data, semi-structured data, and other relevant data, thereby clarifying data security and attaching attribute information labels to the data[5]. In judicial-related fields, the interpretability mechanism of generative artificial intelligence algorithms can improve the credibility of judicial decisions and provide transparent support for risk prediction[6].

2.2 Agile Early Warning Response

The migration of data processing from centralized cloud computing to distributed computing and real-time data analysis methods near data sources or user terminals significantly shortens early warning latency. For example, power enterprises rely on artificial intelligence large models to detect equipment anomalies in real time, reducing fault early warning response time to the second level[7]. In the context of ship safety management, Zhang Jing[8] proposed the "one-network service" platform, which uses data sharing to achieve rapid transmission of risk signals. Sang Xianjun et al.[6] further verified the application potential of large models in the military security field; the training data of these models cover relevant characteristics of risk scenarios, thereby improving the adaptability of early warning systems to various scenarios.

2.3 Closed-Loop Formation of the Prevention System

By using digital technology, it constructs a dynamic model in a virtual environment that completely corresponds to a physical entity. This model accurately represents the entity's operating status through real-time data synchronization and simulation analysis, and it also supports prediction, optimization, and decision-making. Digital twin and reinforcement learning technologies help preventive measures transition from experience-driven to data-driven. For example, laboratory safety management in universities needs to create an "intelligent+" ecosystem, using artificial intelligence to simulate experimental risks and optimize emergency response plans[9]. From the perspective of vocational education, ChatGPT-type technologies enhance students' emergency response capabilities through virtual simulation training, thereby verifying the value of human-machine collaboration in the prevention system[10].

Artificial intelligence technology has triggered a paradigm shift in the field of safety management, driving a value transition from passive response to active intervention. The core of this leap lies in artificial intelligence reshaping the underlying internal logic of safety management, prompting a comprehensive shift from the traditional "experience-driven" model to a "data-driven" model. The deep application of artificial intelligence in safety management will focus on enhancing the breadth and precision of multi-modal data fusion, continuously improving autonomous decision-making capabilities, and vigorously exploring more efficient and transparent human-machine collaborative governance models. Its goal is to build a more mature and universal intelligent safety management practice model, ultimately achieving a comprehensive improvement in safety management effectiveness from local optimization to systemic qualitative change, and providing innovative and implementable safety governance solutions for many fields such as smart cities, critical infrastructure, industrial production, and public safety.

3. Future Path

The future path of artificial intelligence empowering safety management lies in the multi-dimensional integration of technological iteration, scenario expansion, and governance innovation. Its focus is on breaking through current technical bottlenecks and building a more adaptive, inclusive, and forward-looking safety management system. In light of current technological development trends and practical requirements, future research can mainly focus on the following three directions while incorporating innovative research perspectives.

3.1 Technological Integration: The Leap from a Single Tool to Systemic Intelligence

3.1.1 Collaborative Innovation of Multi-Module Data Fusion and Edge Computing

The current problem that data collection relies on a single data source needs to be solved. In the future, it is necessary to build an integrated multi-module perception network, integrating multi-source data such as satellite remote sensing, drone inspection, Internet of Things sensors, and social media public opinion, and to achieve data preprocessing and real-time analysis through edge computing technology. At the same time, it should achieve cross-departmental data sharing under the premise of privacy protection, thereby improving the accuracy and coverage of risk prediction.

3.1.2 Breakthroughs in Autonomous Evolutionary Decision-Making Systems and Causal Reasoning

Existing artificial intelligence models are mostly used for situational analysis but lack causal analysis and reasoning capabilities, making them unable to cope with novel risk scenarios. In the future, it is necessary to adopt the method of "vaccine-style training" to establish a digital twin system that can simulate the evolution paths of multiple risks. Through reinforcement learning, it enables artificial intelligence models to repeatedly simulate attack and defense strategies in a virtual environment, thereby achieving autonomous evolution of defense capabilities. In the field of cybersecurity, it can build an intelligent gaming system (adversarial network) where two neural networks compete with and improve each other to simulate novel virus attack patterns and train the defense system to autonomously generate response strategies. It should introduce causal reasoning computation methods to enable artificial intelligence to identify risk causal chains, shifting from "passive response" to "active intervention at causal nodes," thereby achieving precise application of preventive measures.

3.2 Application Scenarios: From Vertical Fields to the Three-Dimensional Expansion of Complex Systems

3.2.1 Multi-Level Risk Co-Governance in Smart Cities

Smart cities involve diverse systems such as transportation, energy, and medical care. Their safety management needs to break down departmental barriers and establish cross-domain risk collaboration mechanisms. In the field of smart transportation, it can use vehicle-road coordination artificial intelligence systems to perceive risks such as road congestion and abnormal weather in real time, and it can also link with the power grid to adjust street lamp energy consumption and optimize emergency ambulance routes. It should introduce digital twin technology to build an urban risk simulation platform, simulating the transmission of systemic risks in scenarios such as extreme weather and public health events, and planning emergency measures in advance.

3.2.2 Cross-Regional Collaboration in Global Risk Governance

Global risks such as climate change and life safety need to break through national boundaries and establish a cross-border, cross-regional artificial intelligence collaborative governance network. In the relevant work of global epidemic monitoring, it can use artificial intelligence to analyze medical data from multiple countries, determine the transmission patterns of novel viruses, and link with international systems to carry out dynamic control. In the field of marine ecological protection, it uses satellite artificial intelligence monitoring technology to trace the diffusion paths of transnational pollutants and promotes the establishment of an international mechanism featuring "data sharing, shared responsibility, and collaborative governance."

3.3 Governance Model: Paradigm Innovation from Human-Machine Collaboration to Human-Machine Co-Governance

3.3.1 Dynamic Balance between Ethical Governance and Technological Innovation

Ethical issues such as algorithmic bias and data privacy need to be addressed through a "technology-institution" dual-wheel drive. It should develop unbiased and fair algorithms, reducing the impact of data bias on decision-making through technical means. It should improve the ethical guidelines for artificial intelligence safety applications and establish a dynamic regulatory mechanism for algorithm filing and impact assessment. Within the judicial field, human judges need to conduct ethical reviews of decisions made by generative artificial intelligence, ensuring that legal fairness and public order and good morals are not compromised by technology.

3.3.2 Collaborative Improvement of Public Safety Literacy and Artificial Intelligence Governance Capability

It enhances public awareness of artificial intelligence risks, and in the future, it should further strengthen the public's ability to participate in artificial intelligence governance. It should integrate basic artificial intelligence knowledge into primary and secondary school curricula, enabling adolescents to develop algorithmic thinking and data security awareness. For digitally disadvantaged groups such as the elderly, it should develop artificial intelligence safety tools that meet their needs, preventing the digital divide from widening safety inequality.

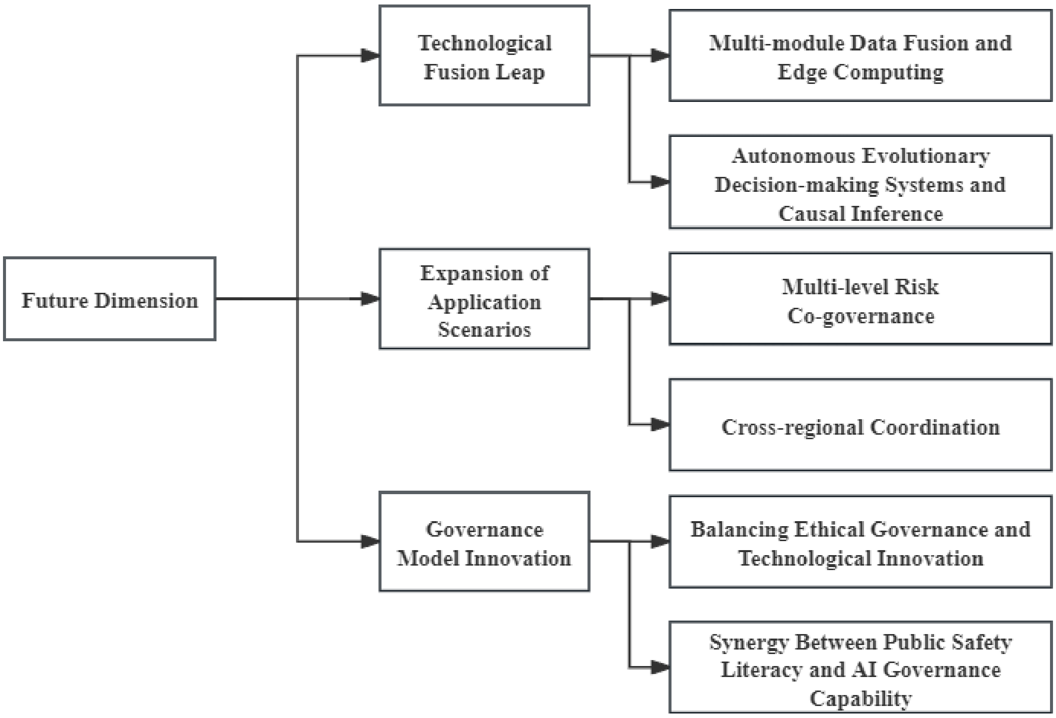

Figure 1 Schematic Diagram of the Three Major Directions of the Future Path

Future research needs to focus on cutting-edge fields such as the cognitive collaboration mechanism of human-machine hybrid intelligence, the capability enhancement of multi-module learning, and cross-domain risk management. At the same time, it should strengthen interdisciplinary research on artificial intelligence ethics, legal regulations, and social governance, injecting both technical rationality and humanistic care into safety management.

4. Practical Approach

4.1 Technological Empowerment: Building a Full-Chain Intelligent Safety Protection System

4.1.1 Construction of an Agile Early Warning Network Supported by Edge Computing

It deploys edge computing devices in critical infrastructure to achieve local data analysis and early warning response. In the field of natural disaster early warning, it can integrate meteorological satellite data, geological sensor data, and social media disaster reports, and use edge nodes to analyze abnormal signals in real time, thereby reducing the time delay from data collection to warning issuance. Power enterprises rely on edge artificial intelligence to monitor in real time the vibration data of transmission lines and the wind loads on wind power equipment. Once an abnormal oscillation is detected, it immediately triggers an alarm and simultaneously pushes location information and possibly faulty component information to the maintenance team.

4.1.2 Closed-Loop Formation of a Digital Twin-Driven Prevention System

It builds digital twins in fields such as industrial manufacturing and urban management, achieving accurate simulation of the operational status of physical entities through real-time synchronized data. University laboratories use digital twin technology to recreate experimental scenarios, and artificial intelligence pre-simulates risks such as chemical leaks, equipment failures, and sudden experimental explosions, thereby deriving optimal emergency plans and delivering them to management personnel.

4.2 Institutional Innovation: Establishing a Rule System for Human-Machine Collaborative Governance

4.2.1 Ethical and Legal Regulations for Artificial Intelligence Safety Applications

For high-risk related fields, such as medical care and judiciary, it formulates mandatory ethical norms that require artificial intelligence decisions to be explainable. For general application scenarios, it establishes industry self-discipline conventions and encourages enterprises to disclose algorithm principles and data sources. In scenarios involving life safety or causing major property losses, such as medical surgeries and nuclear facility control, it requires that artificial intelligence suggestions must be confirmed by human experts before execution. For the power outage decision system in smart grids, when a large-scale blackout procedure is initiated, it requires dispatchers to pass biometric verification.

4.2.2 Improvement of Cross-Departmental Collaborative Governance Mechanism

It establishes a dedicated artificial intelligence safety unit within the emergency management department to coordinate data sharing and operational collaboration among public security, transportation, health, and other departments. For example, during natural disaster emergencies, the rescue resource allocation plan generated by artificial intelligence must undergo joint review by multiple departments to ensure a balance of needs across different domains. For cross-border and cross-basin risks, it promotes the establishment of regional artificial intelligence safety alliances, unifying technical standards and data exchange agreements.

4.3 Humanistic Cultivation: Shaping a Safety Culture Ecosystem of Human-Machine Symbiosis

4.3.1 Stepped Improvement of Public Safety Literacy

It offers general education courses on "Artificial Intelligence and Safety" in primary and secondary schools. Through virtual simulation experiments, such as using ChatGPT to simulate fire evacuation decision-making, it cultivates students' awareness of data security and algorithmic critical thinking. It designs age-friendly artificial intelligence safety products for the elderly. By means of community classes and family pairing, it enables the elderly to master the use of smart devices, thereby avoiding safety problems caused by the digital divide.

4.3.2 Two-Way Construction of Human-Machine Trust Relationships

It adopts a "glass box" algorithm design to explain the key decision-making logic to users, such as the weights of characteristic variables in a health risk prediction model, and it uses a visual interface to enhance users' understanding and trust in artificial intelligence. It establishes a closed-loop system of "user feedback-algorithm optimization," allowing the public to raise objections to artificial intelligence safety decisions and participate in revisions. For traffic signal optimization algorithms in smart cities, it dynamically adjusts parameters based on citizens' travel habit feedback, thereby strengthening the public's sense of identification with the artificial intelligence system.

Leveraging technological innovation, institutional constraints, and humanistic attention, artificial intelligence can truly become the "smart engine" of safety management, enabling a virtuous cycle of prediction accuracy, warning timeliness, and prevention systematicity, and ultimately building a full-coverage, fully intelligent safety protection system.

5. Conclusion and Outlook

This paper focuses on artificial intelligence-enabled safety management and points out that against the backdrop of digital transformation and increasing risk complexity, artificial intelligence promotes safety management from passive response to active defense. Through technologies such as machine learning, edge computing, and digital twins, it achieves precision in prediction, agility in early warning, and the closed-loop formation of a prevention system. At the same time, contradictions such as algorithmic bias and human-machine trust crises exist in technological application. Therefore, this paper proposes a "three-dimensional collaboration" framework that includes formulating ethical guidelines, ensuring human control over key decisions, and promoting public safety literacy education. In the future, artificial intelligence-enabled safety management needs to achieve technological breakthroughs in multi-module data fusion, edge computing collaboration, and autonomous evolutionary decision-making systems, combined with causal reasoning to shift from passive response to active intervention. Its application scenarios should expand from vertical fields to multi-level risk co-governance in smart cities and cross-regional collaboration in global risks. In terms of governance models, it should balance ethics and technological innovation, enhance public safety literacy, and build human-machine trust relationships.

References

[1] Zhao Lei, Zhou Dawei, Pan Peng. Opportunities and Challenges of Artificial Intelligence Empowering Construction Safety Management. Journal of Engineering Management, 2025, 39(02): 100-105.

[2] Xu Qiang, Wang Xin, Wang Jie. Construction of a Community Smart Health Service Platform Based on Deep Big Data Profiling. Digital Communication World, 2025, (03): 76-78.

[3] Zhang Qian, Li Bo. Construction and Application of Engineering Project Risk Assessment and Early Warning Systems. Brick-Tile, 2025, (05): 123-125.

[4] Fan Lili, Zhang Xiang, Ye Zhipeng, Wu Xiaolong, Shen Feiwei, Dong Peng. A Joint Discussion on the Theoretical Explanation and Development Path of Accelerating the Development and Utilization of Public Data Resources. Journal of Chang'an University (Social Science Edition), 2025, 27(01): 1-37.

[5] Zhou Keke. Research on the Data Security Governance Framework and Its Application in Government Data Governance. Thesis Advisor: Ding Hongfa. Guizhou University of Finance and Economics, 2024.

[6] Sang Xianjun. Systematic Construction and Operation of Judicial Algorithm Interpretation Mechanisms. Modernization of Rule of Law, 2025, 9(03): 158-175.

[7] Zhuo Ying, Li Geng. Exploration and Practice of Multi-Region Collaborative Power Communication Safety Management. Electric Safety Technology, 2025, 27(04): 25-28.

[8] Zhang Jing. Research on Safety Management of Shanghai Inland River Ships from the Perspective of Technological Empowerment. Thesis Advisor: Rong Zhi. Shanghai Normal University, 2020.

[9] Wang Fan, Zhang Liucheng, You Tingting, Li Wantao. Analysis and Prospect of Safety Management in University Teaching Laboratories in the Digital Age. Laboratory Science, 2024, 27(04): 204-209.

[10] Guan Yongcai, Du Guobing, Liu Jiali, Yang Jingxin. Research on ChatGPT-Type Technologies Empowering High-Quality Development of Vocational Education. Journal of Xiangyang Vocational and Technical College, 2024, 23(05): 11-16.