Emergency Management
Science and Technology

ARTICLE

Open Access

https://doi.org/10.48130/emst-0026-0002
Emergency Management Science and Technology **2026**, 5: e002

Construction and prospect on the new disciplines by the symbiosis between artificial intelligence and security & safety

Chao Wu[1,2,3*], Bing Wang[1,2,3] and Zhiyong Shi[1,2,3]

[1] *School of Resources and Safety Engineering, Central South University, Changsha, Hunan 410083, China*
[2] *Safety & Security Theory Innovation and Promotion Center, Central South University, Changsha, Hunan 410083, China*
[3] *Safety & Security Science and Emergency Management Center, Central South University, Changsha, Hunan 410083, China*
* Correspondence: wuchao@csu.edu.cn (Wu C)

Abstract

To address the prominent security and safety (SS) problems, arising from the rapid development and widespread applications of artificial intelligence (AI), and to predict the future impact of AI on SS, it is necessary to establish AI–SS disciplines as soon as possible. This article employed methods such as literature review, logical reasoning, new discipline construction, and prediction to conduct an innovative investigation into new disciplines through the symbiosis between AI and SS. After reviewing the typical studies on the SS issues of AI, from its infancy to the current 'explosion period', three-level, and three-category problems of AI–SS are analyzed. Accordingly, three corresponding professional frameworks with a perspective of new disciplines for AI–SS are proposed. Also, the definitions, connotations, research objects, knowledge systems, and interdisciplinary aspects of the three disciplines are elaborated upon. Lastly, the typical research directions, and the symbiotic new disciplines on AI–SS in the future are discussed. The research results indicate that the SS issues associated with the development of AI, will become increasingly severe in the future. The SS issues of AI systems themselves, the applications of AI to improve the quality and efficiency of SS in various fields, and the new SS issues arising from the development and applications of AI, will form three important research and professional fields. Therefore, it is necessary to establish three new disciplines: the Intelligent System SS, SS Intelligence Engineering, and the Intelligent Domain Risk Governance, to adapt to sustainable SS development of AI, and prevent and control various new risks derived from AI and its applications.

Citation: Wu C, Wang B, Shi Z. 2026. Construction and prospect on the new disciplines by the symbiosis between artificial intelligence and security & safety. *Emergency Management Science and Technology* 5: e002 https://doi.org/10.48130/emst-0026-0002

Introduction

Since its initial development, the concept of artificial intelligence (AI) and its research fields has rapidly expanded over many years. The field of AI has covered electronics, information, automation, and other disciplines. With improvements in computer software, hardware, the Internet of Things, and other technologies, AI has made rapid progress in the past decade or two. In this context, in 2018, Carnegie Mellon University (and others), began AI undergraduate programs. In late 2018, the Ministry of Education of China also issued the AI specialty directory, which was added as a secondary discipline under the first-level discipline of electronic information, with the specialty code 080717T. In 2019, many Chinese universities began to enroll students in AI. In December 2024, the Ministry of Education of China also added a number of AI related majors in the updated Professional Catalogue of Vocational Education, including AI data engineering technology.

One of the biggest problems in AI development is security and safety (SS), but the focus of AI development has been on innovation rather than SS, therefore SS often lags behind the research and development of AI technology, and is not given enough attention, which also leads to the current situation where the research of SS intelligence lags behind the development of intelligence. AI had been developing for so many years, and there have been few actions actually focusing on AI security. Until AI made rapid progress in recent years, SS research of AI has become relatively hot. In China, for example, a secondary discipline, named the Intelligent SS, has been set up in the first level discipline of Safety Science and Engineering (code 0837) in the Introduction to Graduate Education Discipline and Basic Requirements for Degree (Trial Version) issued in January 2024. However, there is no major similar to the Intelligent SS in undergraduate education. Even though, in 2025, a proposal on building up new major of training new talents for SS intelligence was put forward[1], the detailed discipline construction on SS AI has still has not been discussed.

AI is well known as being the current, and a future trend, as well as a hot topic for scientific and technological development, and AI SS issues will become ever more prominent. On the basis of summarizing the typical research on AI SS, this paper will focus on the SS issues of AI, and carry out exploration and research from the perspective of new discipline and specialty construction, to provide the solutions of discipline and specialty construction to deal with the SS challenges brought by the current and future development of AI. The goal of this research is to reverse the passive status that the setting of SS disciplines in colleges and universities lags behind the development of high-tech, and meet the demand for new SS professionals to make contributions to the sustainable and safe development of AI and future human security.

Overview of typical research on AI SS

The SS issues of AI have been accompanied by the development process of AI, which has attracted great attention and research investment from researchers and users since its inception. AI has become the forefront of high and new technology, and has been widely used, however AI SS issues are more prominent and important. The following is a brief overview of AI SS research from a time perspective.

SS issues in AI R&D

The development of AI has passed more than half a century since Alan Turing published the landmark paper 'Computing machinery and intelligence'[2] in 1950. He proposed the 'Turing Test' as the criterion to judge whether a machine has intelligence, which became the early theoretical source of the concept of AI. Later, McCarthy first proposed the term AI and established it as a research field[3] at the Dartmouth Conference in 1956, which became a landmark event in the birth of the AI discipline. Afterwards, AI experienced a long period of slow development. It was not until the past decade or two that AI became a shining star of high-tech. In this process, the issues and research of AI and SS have always accompanied each other, evolving and developing. The attack and defense game of AI and SS can be roughly divided into the stages outlined as follows.

The embryonic stage of security defense technology of preset rules (1950–1990)

The early AI system, born in the context of the Cold War, was mainly used in the field of military security, such as password analysis and battlefield situation deduction. The core feature of this stage is that the expert system dominates, and threat identification is realized through manual coding rules. In the initial network security field in the 1990s, network intrusion detection technology was born. Known attack characteristics were identified by preset rules and maintained manually by security experts, showing the application and limitations of early AI in security monitoring.

The data-driven security awareness technology breakthrough period (1990–2010)

The application of statistics theory has promoted the paradigm shift of security defense systems. For example, the algorithm is used for network anomaly detection, discovering unknown threats through traffic feature modeling, and establishing intrusion detection benchmarks, which has spawned a security system based on machine learning; using neural networks to analyze fraud detection in financial transactions; millisecond retrieval of facial recognition and face databases have greatly improved the efficiency of border security inspection; intercepting unknown malware through program behavior analysis, breaking the limitations of traditional feature detection; and Microsoft and other companies realize the automatic mining of software vulnerabilities and the generation of repair suggestions.

The attack and defense game technology explosion period of deep learning (2010–2020)

The rise of deep neural network reshapes the security attack and defense pattern, enabling video monitoring systems to have real-time behavior analysis capabilities, and helping the security technology revolution. It adopts reinforcement learning to simulate network attack and defense, and automatically generates confrontation samples to detect model weaknesses, which greatly shortens the intrusion response time. However, technology evolution is accompanied by risk escalation, and improving data detection efficiency can also be used by attackers to forge attack payloads. For example, Sarker et al.[4] applied AI technology to network security, proposed a defense framework system, emphasized the double-edged sword effect of AI in network security, and called for the construction of a security ecosystem of 'intelligent defense + ethical constraints' to meet the complex challenges of the future digital space.

The security revolution technology fusion period of cognitive intelligence (2020 to present)

The rise of multi-modal large models has pushed security defense into the stage of cognitive intelligence. The initial application of the birth of ChatGPT in security analysis, shows that its ability to process natural language threat intelligence exceeds the traditional rule engine. Integrating the physical platform and AI, it can simulate the chain reaction of nuclear facilities being attacked by UAVs, and realize the security rehearsal of digital twins. The integration of quantum computing and security technology has given birth to new defense paradigms, and promoted the development of quantum cryptography. Encryption algorithm cracking computing technology has also led to the research of anti quantum attack security systems. For example, Johnson[5] believes that meta-cognition is the key mechanism to improve the security of AI systems. By giving AI the ability of self-reflection, it can realize the transformation of security paradigm from passive response to active prevention and control, so that future AI systems will more independently identify risks, optimize decisions, and become an intelligent partner for humans to face complex security challenges.

For more overview of AI's own system security, one can refer to more papers[6,7]. For more than 60 years, the interaction between AI technology evolution and security has followed the law of attack defense spiral development: each defense technology innovation will stimulate the upgrading of attack means, and the confrontation process will give birth to a new defense paradigm. In general, the SS problems in AI R&D mainly belong to the scope of AI specialty.

AI technology applications in various SS fields

For more than half a century, the development of AI technology has attracted great attention and wide application in various fields. However, the time and maturity of the application of AI in SS fields are different. AI technology has been applied relatively early in national security, military security, and information science, and later has been applied in social public security, urban security, and industry. AI has been widely and successfully applied in more SS fields in the past decade or two. With the iterative upgrading of AI technology itself, and its deep application, the concept and mode of SS supervision have also been profoundly changed. In information security, internet security, financial security, power safety, traffic safety, medical and health safety, food safety, monitoring safety, industrial safety, agricultural safety, commercial security, etc., there is a trend of '+ AI', and some new issues are derived from the wide applications of various SS fields by 'fields + AI + new SS'. Some typical applications in various SS fields are summarized in the following sections.

Applications of AI in national and military security

Typical examples are: AI is applied to real-time decision-making of automatic aviation systems, so that unmanned reconnaissance aircraft require minimal manual intervention in flight operations; through network security and intelligence analysis, fast data access is ensured and aided to help military decision-making; AI is applied to electronic warfare and network security to improve the reliability of defense systems.

The following are some examples of specific studies. Chen et al.[8] in 2004, focused on the application of information and SS intelligence science in the field of homeland security emphasized the improvement of security protection capability through the integration of information technology, communication, and transportation systems, and provided key application scenarios in homeland security, providing reference for subsequent related research. In

2005, Chen & Wang[9] discussed the application potential, challenges, and future development direction of AI technology in the field of homeland security by the introduction of special issues, and systematically expounded the key role of AI in homeland security from the perspective of national security strategic needs, which had a forward-looking guiding significance in the application of AI in the field of homeland security at that time. In 2022, Kharazishvill & Kwilinski[10] proposed an AI-based dynamic calculation framework for national security index thresholds, which provides intelligent solutions for security decision-making in complex environments. Through real-time data-driven threshold adjustment, the response speed to threats is improved, and man-made bias is reduced. In 2022, Sanclemente[11] discussed the problem of cognitive bias faced by the application of AI in national security and intelligence analysis, and expounded the systematic risk of AI in cognitive bias from the perspective of systematic solutions.

Applications of AI in the field of social public security

Some typical examples are as follows. Through intelligent monitoring, data analysis, and other means, the public security guarantee ability has been effectively improved; face recognition, voiceprint recognition, and relevant technologies are used to achieve accurate verification of user identity, and to improve the accuracy of identity authentication and system security; through deep learning network training, real-time analysis of network traffic and user behavior, potential security threats are rapidly identified and the construction of intelligent firewall and intrusion detection systems are set up; through intelligent analysis of historical crime data, crime patterns and trends are predicted, and decision-making support for law enforcement agencies are provided. In 2012, Park et al.[12] applied a computer model simulating the behavior of complex systems in the field of counter-terrorism and public security, proposed a new decision support tool integrating distributed computing and group intelligence, and provided a case for public security from data-driven to intelligent collaboration. However, its application is relatively complex, and there are still practical difficulties in computing, privacy protection, human-computer collaboration, etc. In 2014, Vaseashta[13], based on the emerging scientific and technological trend driven by interdisciplinary integration, constructed a forward-looking framework to respond to the rapid evolution of science, technology, and information fields and transform scientific and technological prediction from passive tracking to active design, to provide dynamic decision-making support tools for scientific and technological governance. In 2018, Drăgoicea et al.[14] proposed to build an elastic and customizable public security service system through data intelligence-driven service transformation, and demonstrated the application case against the public security service paradigm. In 2019, Radulov[15] proposed to reconstruct the four pillars of the traditional security system through AI technology in response to the AI driven security paradigm revolution, and described its internal contradictions and future evolution path. It also pointed out that pure technology determinism would lead to systematic risks. In 2020, Nasare et al.[16] proposed multimodal technology integration to build an active security network for womens' security intelligent protection systems, so as to transform passive security protection into active security prevention. At the same time, they built a 'monitoring evaluation response' closed-loop, and solved the contradiction between privacy and security needs, providing a gender sensitive technology paradigm for the field of smart public security. In 2023, Mahor et al.[17] discussed the application of IoT and AI technology in the field of public security, providing new ideas and solutions for improving the level of urban public security, which has certain practical significance.

Applications of AI in the field of urban SS

Cities are the areas with the most concentrated populations and wealth, the main place for human life and production, the gathering place that contains and displays human civilization, and a large and complex system. Its SS is crucial and the top priority in all fields. Therefore, AI has been applied earlier and widely in the field of urban SS. Some specific research examples are as follows. In 2017, Srivastava et al.[18] summarized the application system, technical challenges, and future development direction of AI in smart city security, discussed the interdisciplinary collaboration mechanism including policy makers, technology developers, and sociologists, and provided a preliminary theoretical framework for the application of AI in complex urban system security. In 2018, aiming at the safety framework of an intelligent transportation system based on vehicle road collaboration, Tokody et al.[19] proposed a multi-dimensional safety enhancement scheme by integrating the collaborative design of autonomous vehicles and intelligent infrastructure, and verified using test cases, which explored the transformation of the safety paradigm of intelligent urban transportation system. In 2021, Wang[20] proposed a systematic integration model from theory to practice around the security management framework of security intelligence in the Safety 4.0 era. In 2022, Wang et al.[21] built an intelligent-leading theoretical framework for security management around the frontier issue of security intelligence in the big data environment, and demonstrated its application practice system with urban security management practice cases.

Applications of AI in the field of production safety

Some typical examples are: reducing the accident rate through intelligent monitoring, early risk warning, and other means. Through intelligent analysis of historical accident data, equipment operation parameters, and environmental indicators, dynamic risk prediction models are built to predict various emergencies and improve accuracy. Real-time recognition of behaviors, such as breaking into dangerous areas without a safety helmet, is made based on visual AI. Using voiceprint recognition to analyze abnormal equipment noise, early fault diagnosis is realized. Neural networks are used to predict the remaining life of equipment, improve maintenance efficiency, and reduce maintenance costs. Some specific research cases are as follows: In 2013, Yampolskiy & Fox[22] built a general AI safety engineering framework, which aims to solve the possible risk of runaway caused by super intelligent systems from the system level, and took safety as the core constraint goal of the optimization goal, showing the research direction for inspiring subsequent topics such as AI robustness. In 2019, Wang & Wu[23] built a dynamic integration framework of safety intelligence by systematically answering five core theoretical questions in this field of safety intelligence in safety management, providing a theoretical basis for understanding the nature of safety in the era of intelligence. In 2019, Patriarca et al.[24] proposed to achieve the overall improvement of aviation safety performance through progressive active risk management in accordance with the safety intelligence framework in the field of aviation safety, from avoiding accidents to shaping intrinsic safety.

Applications of AI in disaster prevention and reduction

Some typical applications are through the combination of AI, big data, and IoT, accurate early warning and scientific response to disasters are achieved. Based on meteorological data and high-precision remote sensing data, the impact of rainfall on the flood diversion and discharge capacity of urban waters can be assessed, the ponding situation can be dynamically predicted, and disaster warning information can be released in a timely manner.

Through intelligent and comprehensive analysis of urban settlements, emergency shelters, traffic conditions and other factors, evacuation routes and resettlement plans are dynamically generated and optimized, and emergency command and decision-making capabilities are improved. Specific application examples include: in 2024, Shefer et al.[25] integrated AI into regional security prediction technology, improving the accuracy and practicability of disaster prediction on a technical level.

Applications of AI in the field of public health

Some typical applications of AI technology are to help disease monitoring, drug R&D, and epidemic prediction, and improve the ability to respond to public health events. The image recognition algorithm is used to quickly recognize medical image features such as CT's and X-ray's to obtain diagnostic results. Through intelligent analysis of epidemic data, the trend of epidemic development can be predicted, and a scientific basis for epidemic prevention and control can be provided. In 2025, Gwala[26] carried out a comprehensive overview of the application of blockchain technology and AI in cryptocurrency and medical technology, and looked forward to the future development trends, providing valuable reference and enlightenment for researchers and practitioners in related fields.

The current trends on AI SS investigations worldwide can be seen from the literature statistics obtained from the Web of Science and CNKI (the largest data base of Chinese language research documents, Fig. 1). From the trend of data and curve potential in Fig. 1, it can be seen that the number of documents on AI SS has increased sharply in recent years, proving that AI SS will be increasingly important in the future.

Construction of new disciplines by the symbiosis of AI and SS

It can be seen from the above overview that, in recent decades, with the development and popularization of AI technology, AI's own security issues, and its application in the security field are of concern, and remarkable achievements have been made. However, the research on AI SS issues generally lags behind AI technology, and the research on disciplines and professional issues related to AI SS is extremely rare, let alone even on the agenda. However, from the current trend of AI's rapid development, it can be predicted that, AI SS will become one of the most popular research fields in the AI era, and a large number of professionals will also be needed (Fig. 1). The following sections will start with the analysis of the relationship between AI and SS, and then the disciplinary framework will be constructed by the symbiosis of AI and SS.

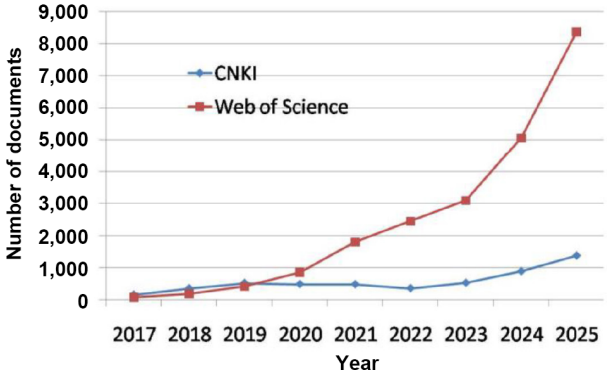

Fig. 1 Literature number search statistics on AI SS from the Web of Science and CNKI using the topic 'AI SS'.

The core foundation of the symbiosis of AI and SS is that both sides need to complement each other in content, paradigms, and problems solving. For example: (1) The complexity of AI has generated new security risks, forcing security research to shift from protecting external attacks to building credible AI; at the same time, the massive threat data and dynamic confrontation scenarios in the security field provide AI with a key training environment and evolutionary goals, driving it to a more robust and interpretable direction; (2) The deep integration of the classic paradigms of traditional information security, such as attack and defense, risk management, and the core paradigm of AI has led to the emergence of new paradigms of intelligent security. Security engineering can use AI for active prediction and adaptive response, and AI systems must be embedded with security design to ensure their reliable deployment; (3) The major challenge facing an intelligent society is essentially a complex of AI problems and security problems, which cannot be solved by a single discipline. This forces the two disciplines to form a new community, jointly define new problems, and jointly build a multi dimensional solution from technology to governance.

Analysis of three-level and three-category problems of AI and SS

AI and SS have been confused in the past, and few people have made in-depth classification. From a professional perspective, AI and SS issues can be divided into three levels and three categories:

The first level is the SS problem of the AI system itself. Its research scope involves the hardware, software, endogenous, and external system SS problems of AI technology, such as the reliability of smart chip hardware, the prevention of the human-computer interaction interface being hijacked and tampered with, the detection of training data being polluted, the defense against external attacks, real-time anomaly monitoring, systematic collapse, and other SS problems. The core content and focus of its research are the reliability, resilience, and self-defense capabilities of the intelligent system itself. This level of SS issue is called the first category of AI SS issues.

The second level is how to apply AI to improve and solve multiple SS problems in various fields, without generating new SS problems. Its scope of applying research is very broad, covering all existing SS fields, such as the intelligence of SS in various fields, the intelligence of risk prediction, the intelligence of accident prevention, the intelligence of security emergencies, etc. Its core content is to create a new pattern of intelligence driving SS in various fields, with the focus on promoting and applying advanced intelligent technology to various SS fields to achieve the SS intelligence. This level of SS issue is called the second category of AI SS issues.

The third level is the control and governance of new SS issues arising in the process of AI R&D and applications of AI. Its research scope covers AI's social and technological risks, AI's ethical conflicts, new AI crime prevention and control, AI fraud prevention and control, AI applications and users' safety management, AI future risk prediction and control, etc. Its core content is to establish a risk management and control system and governance pattern for intelligent society, and balance AI technology innovation and applications with social sustainable SS, stability, and future SS through system design and various regulatory and governance practices. This level of SS issue is called the third category of AI SS issues.

Obviously, in the time dimension, the first category of AI SS problems in the above three levels appeared first, followed by the second and third categories. The three categories of SS problems are not completely progressive in the time dimension. In practice,

the three categories of AI SS problems are interrelated and interactive, and have some degree of intersection. With the application of AI in various SS fields from trial to maturity, the three categories of AI SS issues can also form a spiral upward trend and mode. The three-level and three-category problems of AI and SS and their interrelationships are shown in Fig. 2.

Building new disciplines based on three-level and three-category problems of AI and SS

According to the three-level and three-category problems of AI and SS issues shown in Fig. 2, and the creation paradigm of new disciplines[27], as well as the experiences of the newly built disciplines[28,29], the following three new disciplines or majors can be established:

Intelligent system security and safety

This is a discipline corresponding to the first category of AI SS problems, which mainly studies the SS problems of intelligent systems themselves. This discipline is mainly aimed at the internal and external threats to intelligent systems, and the SS threats to their systems to ensure the SS of intelligent system hardware and software and their systems, focusing on AI algorithm reliability, data security, system robustness, and anti-interference ability, solving security problems such as algorithm vulnerabilities, model pollution, and resistance to attacks, as well as building systems with self verification, fault tolerance, and repair capabilities, and resistance to attacks, so as to ensure the safe and stable operation of intelligent systems in complex environments.

Security and safety intelligence engineering

This is a discipline corresponding to the second category of AI SS problems. It mainly studies the engineering practice of AI applications in various SS fields, enabling AI to empower SS technology, SS engineering, SS management, SS education, SS culture, etc., in various fields, so as to improve the SS level and SS efficiency in various fields, and also prevent secondary SS problems in the process of AI applications. Typical application examples include industry safety intelligence, public security intelligence, health supervision intelligence, natural disaster prediction and early warning intelligence, urban SS intelligence, emergency management intelligence, etc., thus comprehensively upgrading the traditional SS level and SS efficiency.

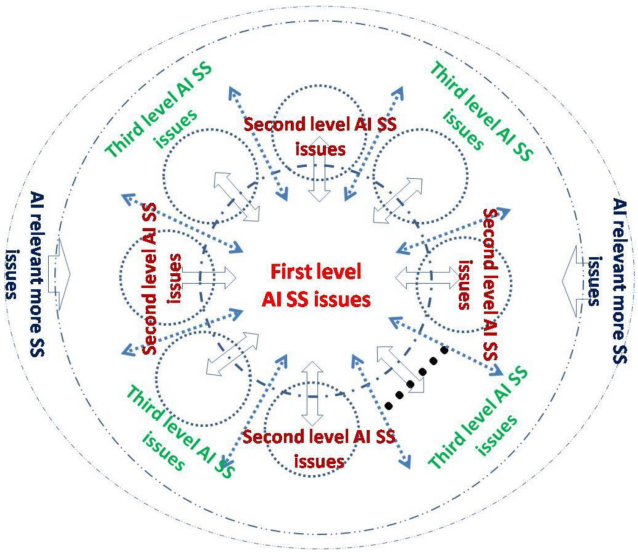

Fig. 2 Schematic diagram of the interrelationships of three-level and three-category problems of AI and SS.

Intelligent domain risk governance

This is a discipline corresponding to the third category of AI SS issues, which mainly studies the secondary SS risks caused by the application of intelligent technology, including technical risks, social risks, economic risks, cultural risks, human SS risks caused by AI, and specific risks such as the new risks of privacy disclosure, algorithm discrimination, and deep forgery caused by AI in various industries and departments. An important mission of smart risk governance is to establish a multi-dimensional governance framework system covering ethical review, legal regulation, and technical prevention and control, and form a new social SS paradigm of human-machine collaboration in times of intelligence.

Theoretical foundation of three new disciplines

The theoretical foundation of the three new disciplines is described below.

The Intelligent System SS is deeply integrated with the principles of AI, complex system science, and active defense theory, etc. Its core is to ensure the credibility of the intelligent algorithm's own decision-making, the manageability of human-computer interaction, and the overall controllability of the system in the event of attack or failure. The essential difference between the Intelligent System SS, and the Information Security of related discipline is that the Information Security mainly focuses on the protection of external threats against data and networks, which belongs to external defense; the Intelligent System SS focuses on the endogenous risks of agents due to design defects, data bias or logical black boxes, and is committed to building the self immunity and self-healing ability of intelligent systems.

The theoretical foundation of the SS Intelligence Engineering is intelligent safety ergonomics and dynamic risk perception theory, etc. It does not simply use AI tools in the security field, but deeply embeds AI into security engineering to realize the whole process intelligence from risk identification, to assessment and control. The difference between the SS Intelligence Engineering and traditional Safety Engineering is that the former emphasizes the realization of real-time dynamic perception, predictive early warning, and adaptive response of safety risks through technologies such as the Internet of Things, digital twins and agents, and the construction of an intelligent safety protection system that can independently evolve and constantly learn.

The theoretical foundation of the Intelligent Domain Risk Governance derives from cognitive domain security management, polycentric governance theory and technological sociology, etc. It mainly studies new social and political risks and their governance caused by information manipulation, algorithm bias, group cognitive polarization, etc., in a digital and intelligent environment. Compared with the traditional Public Management specialty, the Intelligent Domain Risk Governance expands the governance object to the virtual cognitive space and information ecology, etc. Its core challenge is how to coordinate multiple subjects such as the government, technology companies, industry organizations, and the public to jointly deal with new social risks with high concealment and diffusion, such as deep forgery, algorithm manipulation, large-scale cognitive warfare, etc.

These three new disciplines represent a new direction to meet the challenges of the intelligent era: the Intelligent System SS is committed to making intelligent systems more reliable; the SS Intelligence Engineering is committed to making the protection system more

active and intelligent with intelligent technology; and the Intelligent Domain Risk Governance is committed to managing new cognitive and political risks related to social stability caused by the wide application of intelligent technology. They form a future oriented, three-dimensional intelligent security discipline group from the three-category problems of technology core, engineering methods, and social governance.

The distinctions of the three new disciplines

The distinctions of the three new disciplines are as follows. The Intelligent System SS deeply integrates the principles of AI, complex system science, and active defense theory. Its core is to ensure the credibility of the intelligent algorithm's own decision-making, the manageability of human-computer interaction, and the overall controllability of the system in the event of attack or failure. The theoretical foundation of the SS Intelligence Engineering is intelligent safety ergonomics and dynamic risk perception theory, etc. It does not simply use AI tools in the security field, but deeply embeds AI into security engineering to realize the whole process intelligence from risk identification, assessment, to control. The Intelligent Domain Risk Governance derives from cognitive domain security management, polycentric governance theory, and technological sociology. It mainly studies new social and political risks and their governance caused by information manipulation, algorithm bias, group cognitive polarization, etc. in a digital and intelligent environment.

The symbiotic relationship between the Intelligent System SS and the SS Intelligence Engineering is embodied in three levels. In the technical core category corresponding to the Intelligent System SS, it is necessary to ensure the security and credibility of the AI model itself; at the engineering system category corresponding to SS Intelligence Engineering, it is necessary to use AI to build a dynamic defense system. Since this is a large category of majors and all of them are interdisciplinary, there is some overlap, but the emphasis is different. The essence of this symbiotic relationship is a positive feedback cycle: the evolution of AI constantly puts forward new security issues, and the upgrading of security requirements continues to feed back AI technology breakthroughs. It marks the upgrade from an AI-enabling security tool relationship to AI and security together, defining the symbiotic relationship of the future intelligent society, and finally realizing the unification of safe intelligence and intelligent security.

Knowledge structure systems of the three new disciplines

In this section, the research object, category, discipline basis, and external connection with the social system of the three new disciplines are described respectively.

Knowledge structure system of the Intelligent System SS

The purpose of the Intelligent System SS is to cultivate professionals who can ensure the safety, credibility and reliability of the AI system itself. The core is to solve the endogenous risks caused by AI models and data. The professional goal and orientation is to focus on the security of intelligent systems, and ensure their full life cycle reliability from design, training to deployment, and operation. The Intelligent System SS focuses on the SS of the agent itself, and its knowledge structure system and its relationship are shown in Fig. 3.

The following sections will provide a more detailed analysis of Fig. 3.

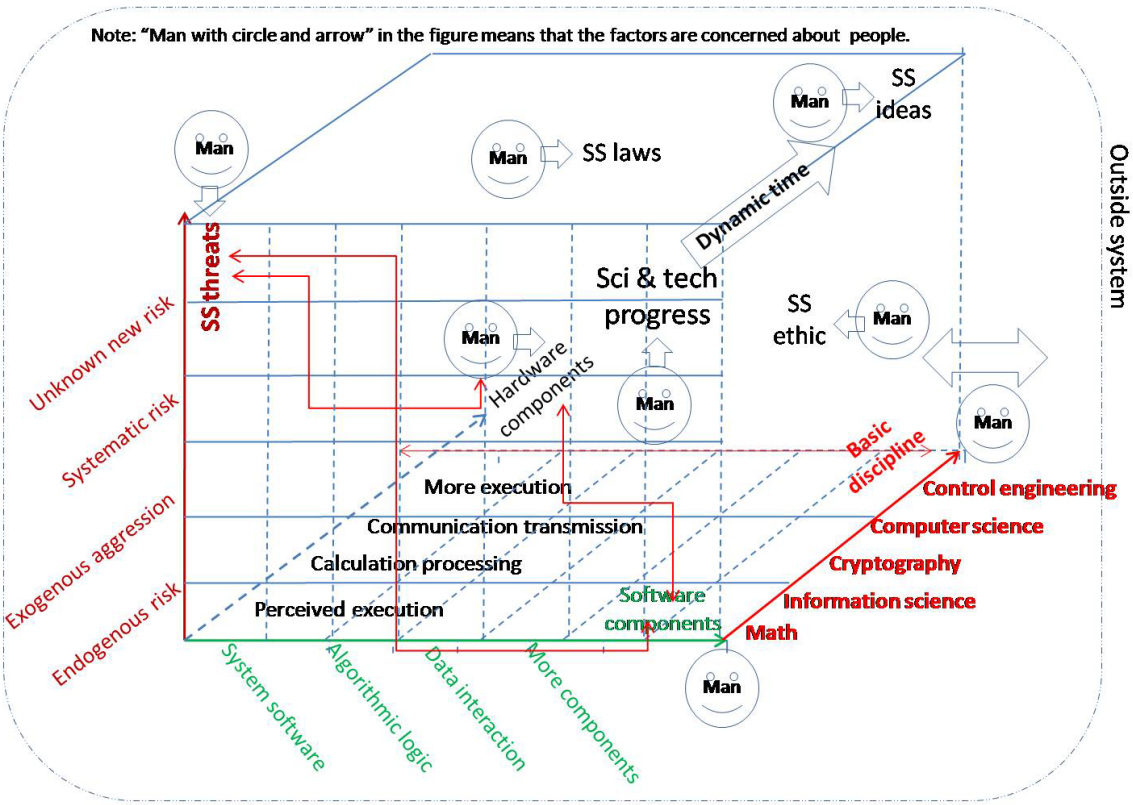

Fig. 3 Knowledge structure system of the Intelligence System SS and its interrelationships.

Software component dimension

The important component of an intelligent system is software, including system software which is related to SS issues such as operating system kernel security mechanisms, virtualization platform protection systems, runtime security protection, trusted verification modules, etc.; algorithm logic which is related to SS issues such as machine learning model reliability, decision algorithm correctness, automatic control process verification, privacy protection mechanisms, etc.; and the data interaction layer which is related to SS issues such as multimodal data cleaning, data processing integrity protection, access control security policy, consistency assurance, etc.

Hardware component dimension

The hardware components of an intelligent system include: sensing execution elements, such as sensors, data acquisition circuits, execution physical mechanisms, authentication elements, security computing startup elements, etc.; computing processors, such as chips, chip installation architecture, memory protection elements, cryptographic coprocessors, etc.; communication transmission elements, such as physical layer signal elements, bus detection elements, channel anti eavesdropping technology, sensitive network integrity protection, etc.

SS threat dimension

The SS threats of an intelligent system include endogenous risks, such as software and hardware design defects, algorithm decision path vulnerabilities, timing channel failures, multi component coupling failures, etc.; exogenous attacks, such as input attacks, physical environment disturbance attacks, supply chain pollution attacks, protocol cracking attacks, etc.; systematic risk domains, such as human-computer interaction trust crisis, risk of autonomous evolution out of control, threat of emergence of swarm intelligence, cross-border data sovereignty conflict, etc.

Discipline-based dimension

The discipline basis of the Intelligent System SS mainly includes mathematics, informatics, cryptography, computer science, control engineering, etc.

It can also be seen from Fig. 3 that the components and elements of the above four dimensions are interrelated and affect each other, forming a dynamic operating system. Each component and element of the system are related to human factors, subject to SS laws and regulations and ethics, and constantly produce dynamic exchange of input and output with external systems. In addition, in the current stage when human beings can play a leading role in intelligent systems, it is critical that forward-looking advanced SS concepts lead the design of all components and elements of intelligent systems, and the operation of the system.

From Fig. 3, the core courses of the Intelligent System SS may include advanced machine learning, optimization theory, introduction to trusted AI, adversarial machine learning, privacy computing, AI model security testing and verification, intelligent system security architecture, AI chip and hardware security, etc. A detailed curriculum may depend upon the real situation of colleges and universities.

Knowledge structure system of the SS Intelligence Engineering

The purpose of the SS Intelligence Engineering is to train engineers who can use AI technology to build an active, adaptive, and evolutionary next generation SS protection system. The core is intelligence driven SS. Professional goals and positioning go beyond the static and post response modes of traditional SS engineering, and realize real-time perception, intelligent prediction, and the automatic disposal of security risks.

SS Intelligence Engineering focuses on the application of AI to ensure and improve the SS level and work performance in various fields. The knowledge structure system of this discipline is shown in Fig. 4. The research scope and application scenarios for realizing SS intellectualization mainly include: the application of AI in national security, military security, social public security, prevention and reduction of natural disasters, prevention and control of sudden epidemic, accident disaster prevention and control, SS supervision and management, enterprise safety production management, SS emergency assistance, safety science and technology research and development, information security assurance, traffic safety, urban SS, SS talent training, safety education, safety media, etc. The technical support of this discipline requires intelligent technology, industry background technology, market development ability, technical and economic analysis ability, engineering practice experiences, etc. The discipline foundation mainly includes intelligent science, SS science, management science, social science, etc. It can also be seen from Fig. 4 that the components and elements of SS Intelligence Engineering are also interrelated and affect each other, and each component and element is related to human factors. They are also constrained by SS laws, regulations and ethics, and constantly generate dynamic exchanges of input and output with external systems. In addition, the application of intelligent technology in various SS fields also needs correct SS concepts to guide it. Because the background and technology of each application field are very different, and there are many scenarios, SS Intelligence Engineering needs to be integrated into different professional modules, which may form a large sub discipline group of SS Intelligence Engineering.

From Fig. 4, it can be deduced that the core courses of the SS Intelligence Engineering include security big data engineering, data analysis, digital twin technology, security situation awareness and prediction, anomaly detection algorithm, intelligent traceability, intelligent security operation design, emergency response automation, industrial Internet security engineering, etc. Detailed curriculum systems may depend upon the real situations of colleges and universities.

Knowledge structure system of the Intelligence Domain Risk Governance

The Intelligent Domain Risk Governance aims to cultivate compound governance talents who can identify, evaluate, and govern social, cognitive, and ethical macro risks caused by disruptive technologies such as AI. Professional goals and positioning are to deal with new types of intellectual domain risks such as algorithm manipulation, cognitive warfare, and deep forgery, and to build a multi coordinated governance ecosystem. The knowledge structure system of this discipline is shown in Fig. 5.

The Intelligence Domain Risk Governance mainly focuses on SS supervision and governance in the process of intelligent technology R&D and its wide applications. The research scope mainly involves technical risks, social risks, economic risks, human future risks, etc. generated by AI, and carries out research on SS management of AI hardware, SS management of AI software, SS management of AI system, SS governance from local to overall, SS governance of secondary SS problems in the application fields from short-term to long-term. Specific examples include: SS norms and standards of AI technology research and development, SS management norms and standards of AI applications, human-computer ethical conflict, responsibility attribution of AI medical misdiagnosis, traceability of

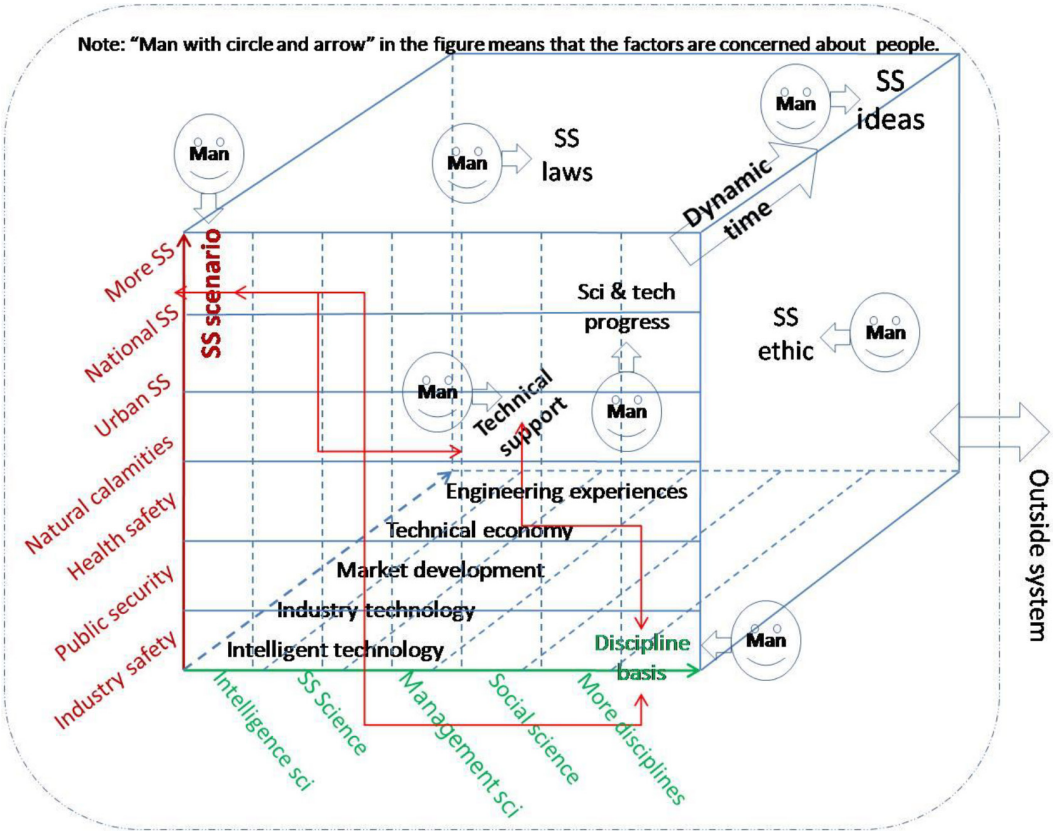

Fig. 4 Knowledge structure system of Intelligence Engineering and its interrelationships.

AI forged contents, AI-assisted network crime supervision, etc.[30,31]. The discipline basis of the Intelligence Domain Risk Governance includes intelligence science, safety science, social science, management science, complex science, etc. The scale of the research object of the Intelligence Domain Risk Governance is relatively large. The main purpose of this discipline is to prevent, control, and balance the risks brought to society by AI technology innovation and applications, and to ensure the SS of human future through institutional design, management, and governance from meso to macro. It can also be seen from Fig. 5 that the components and elements of the Intelligence Domain Risk Governance are also interrelated and interactive, and they are all related to human factors. They are also constrained by SS laws, regulations, and ethics, and constantly generate dynamic exchanges of input and output with external systems. In addition, the intelligence domain risk governance process needs correct SS ideas for guidance.

From Fig. 5, it can be deduced that the core courses of the Intelligence Domain Risk Governance include science and technology sociology, risk social governance theory, AI ethics and governance framework, social computing, computational communication, cognitive psychology and information manipulation analysis, algorithm audit methods and practices, platform governance and content review, science and technology policy design and evaluation, transnational digital governance and cooperation, crisis communication, etc. A detailed curriculum system may be dependant on the real situation of colleges and universities.

Comparison of the characteristics of the three new disciplines

The core quality focus of the major of the Intelligent System SS is to deeply understand the endogenous safety attributes of intelligent systems, and establish engineering ethics; professional ability is the core theory and technology for the system to master the evaluation and enhancement of AI model security. It can design and implement effective protection and verification schemes against algorithm vulnerabilities, data poisoning, model theft, and other threats. The core accomplishment of the SS Intelligence Engineering specialty is to establish a dynamic, active, and adaptive safety protection concept of the next generation, and understand the risk transmission law in the complex human-computer object fusion system; professional ability is to be proficient in using big data, AI, digital twins, and other technologies to conduct security situation awareness, threat prediction, automatic response, and system resilience design. The core quality of the Intelligent Domain Risk Governance major is to have a scientific and ethical perspective, and to allow insight into the social generation mechanism and political impact of new intelligence risks such as algorithm recommendation, deep forgery, cognitive manipulation, etc; professional ability is to master algorithm audit, policy analysis, stakeholder coordination, and multi-level governance tools, and be able to design and promote the implementation of an effective risk governance framework.

More focus and characteristics of the three new disciplines of the Intelligent System SS, SS Intelligence Engineering, and Intelligence Domain Risk Governance, and their relationship are shown in Table 1.

In summary, the interrelationship between the three new disciplines is that new intelligent application fields can be found, and new SS threats can be exposed in the practice of the SS Intelligence Engineering. These results can promote the development of new SS intelligent products, and new intelligent protection technologies in the fields of the Intelligence System SS. The Intelligence Domain

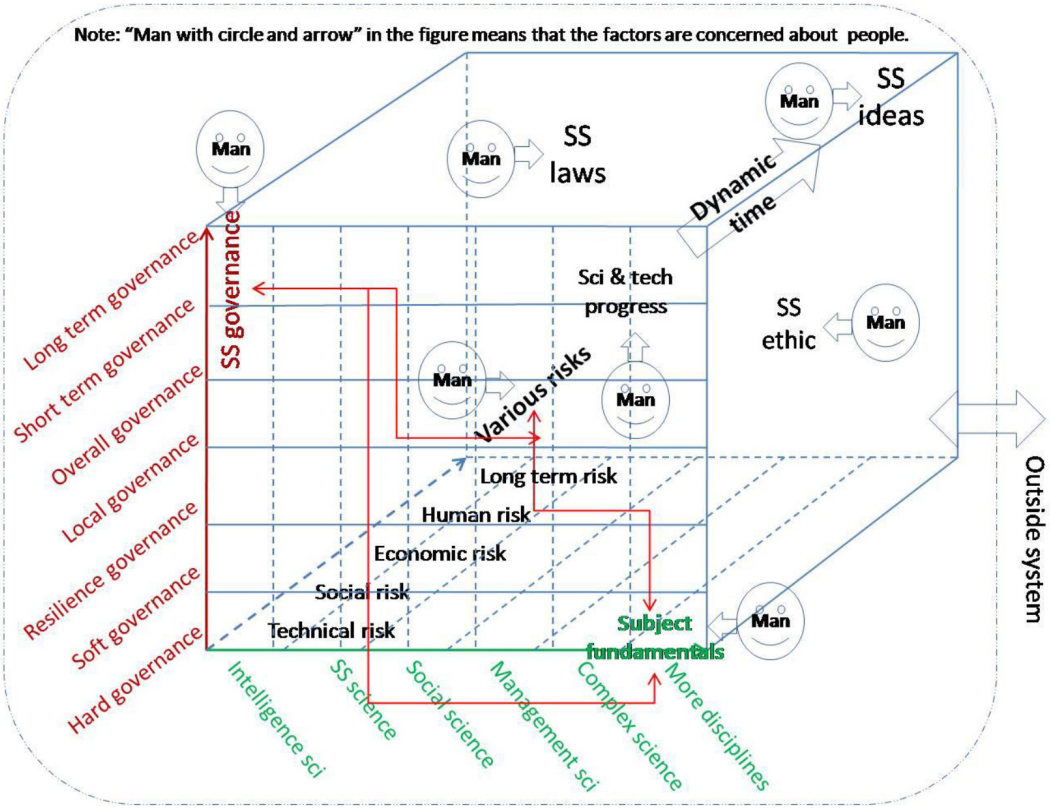

Fig. 5 Knowledge structure system of the Intelligence Domain Risk Governance and its interrelationships.

Table 1. Focus and characteristics of the Intelligent system SS, the SS Intelligence Engineering, and the Intelligence Domain Risk Governance.

Disciplines	Specialty of Intelligent System SS	Specialty of SS Intelligence Engineering[34,35]	Specialty of Intelligence Domain Risk Governance[35–37]
Many existing disciplines and majors belong to (undergraduate)	Such as electronic information, computer science, electronic automation	Such as safety science and engineering, public security technology, cross disciplines of engineering	Such as safety science and engineering, public management, management science and engineering
Focus of R&D	SS problems of the intelligence system itself[32,33]	Intelligent technology is used to solve the SS problems in various fields	New SS problems arising from R&D of intelligent technology and its application process
Action level and scale	SS of intelligent technology ontology layer, SS issues in micro area	Multiple SS application scenarios of social layer, SS issues from micro to meso	Secondary SS of appearance layer, SS issues from meso to macro
The time dimension and SS effectiveness	SS problems appear first (one-way reduction, negative risks)	SS problems appear after the applications (positive and negative two-way, positive effects mainly, negative effects secondary)	SS problems appear relatively late (one-way reduction, negative effects)
The interrelationship of the three disciplines	The specialty provides SS technology base for intelligence systems, and is the core components of SS Intelligence Engineering	Intelligence system achievements are transformed to engineering	Prevention of secondary SS problems arising from intelligent technology R&D and applications in the SS fields
Protection object difference	AI systems are the protection object	SS of various scenarios is the action object	Human social SS is the protection object
The personnel of three disciplines	Mainly AI professionals, supplemented by occupational SS professionals	Occupational SS professionals, supplemented by AI professionals	occupational SS professionals and managers, and supplemented by AI professionals
Examples of collaboration among three disciplines (taking on-board AI as an example)[34–37]	Design on-board AI technology to be safe and reliable, enabling vehicles to run normally and resist external deception and attacks on sensors; the safety upgrading of on-board AI; etc.	Selection of appropriate on-board AI and providing AI design professionals with on-board AI application experience and design philosophy; the maintenance of on-board AI implementation process and environmental safety; ensuring the safety matching between on-board AI and vehicles; etc.	Providing SS management specifications and standards for the design and application of on-board AI technology; formulating insurance terms, market norms for on-board AI failures; Intelligent System SS personnel and SS Intelligence Engineering personnel participate in the reasonable setting and application scope of technical and application indicators in the safety standards and specifications of on-board AI

Risk Governance restricts the abuse and secondary risks of intelligent technology by studying and formulating new rules, and promotes the iterative upgrading of the Intelligence System SS and SS Intelligent Engineering and the improvement of system defense capabilities. The above process forms a triple spiral structure of 'AI technology reinforcement and upgrading—institutional constraints, AI secondary risks—social SS capabilities are enhanced', which essentially constitutes a positive cycle of SS level enhancement in the intelligent era. These three new disciplines together constitute the trinity framework of SS research and applications in the

intelligent era, which not only prevents the risks of intelligent technology itself, but also uses intelligent technology to improve the overall SS level of society, forming a complete discipline system of intelligent SS that spirals and complements each other.

Discussion on the challenges of the three new discipline formation

In future practice, the three new disciplines or majors will obviously encounter many challenges. The three new disciplines or majors aim to meet the security challenges of different fields in the intelligent era. However, their formation and development certainly face deep challenges from the theory, technology, and education system, etc.

The Intelligent System SS is to realize the endogenous security of algorithms, etc. The core challenge lies in the inherent characteristics of intelligent algorithms, such as data-driven, uncertain calculation, and model inference, which are difficult to interpret. The failure of traditional methods based on certainty and completeness verification must break through fundamental scientific problems, such as the determination of the confidence region of the uncertainty algorithm, and transparent monitoring of the black box model in theory and technology. For example, although it is a promising path to transform unknown risks into probability-controllable problems, its large-scale engineering implementation still faces challenges.

The SS Intelligence Engineering focuses on complex systems with human-computer integration. When intelligent systems participate in perception and decision-making, people and machines may have structural deviations in their understanding and judgment of risks. For example, in automatic driving and other scenarios, failure often results from inconsistent cognition, mismatched trust, or fuzzy responsibility boundaries between humans and machines. Therefore, the key of this discipline or major is to establish an effective human-computer two-way understanding and collaboration mechanism, which goes beyond the scope of pure technology, and involves multiple dimensions such as cognitive science and human factor engineering. There is a practice gap of dynamic collaboration.

The Intelligent Domain Risk Governance is to deal with systemic risks at the social field. The challenge is that the development of governance rules (laws, ethics, standards) lags far behind the evolution of AI technology. In the face of new intellectual domain risks such as deep forgery, information cocoon room, and cognitive manipulation, the core challenge is how to establish an effective measurement assessment enhancement technology system, and collaborative governance framework for such social risks that are difficult to quantify.

Since these three new disciplines or majors are the blueprint at present and they are the future AI SS fields, it needs to go beyond a proposing program and forward thinking to transform them into reality. Its success also depends on whether it can break through the theoretical bottleneck of AI credible verification in science, establish a new paradigm of human-computer mutual trust and collaboration in engineering, build an agile and forward-looking rule system in governance, and finally completely break the disciplinary and organizational barriers in education to achieve the real cultivation of innovative talents.

Prospects for the three new disciplines

Some important research directions and possible new branches of the above three new disciplines are outlined in the following sections.

General research directions of the three new disciplines

It can be seen from Figs. 2–5 that the common research directions of the three disciplines are SS concepts, laws, ethics, dynamics, etc. All components and elements involve human factors, and form a complex system. In addition, the goals of the three new disciplines are all for SS. From the above, some important general research directions of these three new disciplines can be inferred and forecasted[38].

The philosophy theories of the sustainable safe development of AI

This is a basic research field integrating with AI technological innovation, environmental responsibility, long-term social development, and human future SS. Its core is to ensure ecological sustainability, ethical inclusiveness, global collaboration, and the safe survival of human beings while improving SS efficiency through intelligent technology. The core forward-looking elements involved include preventive intelligent defense, green technology ethics, global SS resilience, dynamic balance paradigm, human future SS, etc. Its ultimate value is to lead and balance AI technology development, achieve SS through intelligence, feed sustainability through SS, and finally form a civilized development cycle of technology for the good.

Ethical research on the sustainable and safe development of AI

This is a comprehensive field integrating AI technology innovation, social responsibility, and ethics, etc. Its core is to guide the sustainable development of intelligent technology through an ethical framework, and ensure the SS, fairness, transparency, and environmental friendliness of AI technology applications. Its research is the intersection of technology, SS, ethics, society, etc. Its goal is to avoid technology getting out of control or abused. Technology applications do not threaten human SS, economic stability, or ecological balance. The behavior of technology developers and users should prevent algorithm bias, privacy violations, and other issues, form a global AI ethical consensus, and ensure that the development of intelligent technology does not deviate from the original intention of improving human welfare.

Basic theoretical research on the SS development of AI

This is a research field focusing on the underlying logic and scientific basis of building an SS development paradigm of an intelligent system. It aims to reveal the essential laws, core contradictions, and regulatory mechanisms of intelligent technology SS development through systematic and interdisciplinary exploration, such as AI SS complexity research, AI SS cybernetics, and SS information theory, AI computing SS sociology, AI SS technology evolution dynamics, AI risk entropy growth law, AI SS threshold, AI SS key model, AI cross civilization SS axiom, AI SS education system, etc., and form basic theoretical system of AI's own SS by integrating multi-disciplinary basic knowledge supporting AI, such as SS science, computer science, informatics, ethics, sociology, culture, anthropology, etc.

Systematic science on the SS development in AI

This is an interdisciplinary field integrating system science, SS science, computer science, sociology, ethics, management science, etc. The research emphasizes the overall analysis of the interaction between SS development and intelligent technology, reveals the dynamic relationship between AI technology, SS, ethics, environment, and human beings through feedback mechanisms, causal relationship models, etc., and sets a SS threshold for intelligent

technology through multidisciplinary integration and dynamic system analysis, so as to ensure that it serves the ultimate goal of sustainable development.

Futurology on the SS development of AI

This is an interdisciplinary frontier field, which aims to explore the potential path, risks, and opportunities of intelligent technology development, and build a future oriented SS development framework by combining AI, big data, system science, risk management, ethics, futurology, and other disciplines. Its connotation includes the ideas of balanced SS and development, the dynamic SS concept, preventive thinking, and resilient and sustainable development. Research examples include AI technology evolution prediction, future risk scenario simulation, future ethics and governance framework, human-computer collaborative future, etc. The futurology of SS development intelligence is essentially an unknown oriented governance science. Its goal is not only to predict the future, but also to guide technological development to a sustainable, inclusive, and consistent track with the overall interests of mankind through active design. The development of this field will lead the direction of human SS civilization in the 21st century.

Research directions of the Intelligent System SS

Apart from the description above, from Fig. 3, we can also see more forward research directions, and the future development of the Intelligence System SS.

AI hardware system security technology

This direction is mainly aimed at the risk control inherent in AI hardware, such as the research and development of intelligent risk perception components, intelligent multimodal threat identifier, fault causal reasoning decision generator, emergency response controller, trusted link verification technology, dynamic defense technology, elastic recovery technology, countermeasures technology, self-healing network architecture, etc.

AI software system security technology

This direction is mainly aimed at the prevention and control of AI software endogenous risks, such as the enhancement of multi intelligent software vulnerability, the deep neural network against the sample generation mechanism, the immunity of data pollution and information feedback misleading, the risk of parameter reverse execution, code security analysis procedures, infrastructure digital immune system, intelligent deduction of chaotic security scenarios, intelligent prediction of security situation, data trajectory tracking, trusted audit of blockchain data flows, privacy computing attack and defense, consciousness space protection, and the development of brain computer interface data flow firewall.

The AI system attack-defense confrontation game

This direction is mainly aimed at the prevention and control of exogenous risks of the system, such as the expansion of cognitive security dimensions, brain computer interface signal hijacking protection, digital twin cognitive deception defense, psychological manipulation confrontation, group intelligence cognitive deviation correction mechanism, attack and defense intelligence, vulnerability prediction, multi-dimensional attack autonomous perception technology, defense strategy optimization technology research and development.

AI system SS engineering

This direction is mainly aimed at AI system SS issues, such as intelligent defense architecture, autonomous evolutionary defense network, dark network space active trapping system, adaptive authentication protocol in dynamic environment, deep intelligent learning capability, anti risk resilience system, complex SS system theory, SS agent cognitive architecture, human-machine SS cooperation system, risk propagation model, multi-agent game defense dynamics, etc.

After continuous enrichment and consolidation, Intelligent System SS should become a new branch in intelligent science and technology, and can derive more related discipline groups, such as attack and defense game intelligent dynamics, digital twin SS engineering, AI computing SS, consciousness information security, cross modal interaction SS, data security engineering, multi-agent collaborative defense, intelligent emergency technology, intelligent IoT SS technology, etc.

Research directions of the SS Intelligence Engineering

The SS Intelligence Engineering is completely suitable for being a new branch of intelligent science and technology and SS science. From the knowledge structure system in Fig. 4, SS Intelligence Engineering has a great number of application scenarios in various SS fields, such as national security, military security, social public security, disaster prevention and reduction, sudden epidemic, accident disasters, SS supervision, enterprise safety production management, emergency rescue, SS science and technology R&D, information security assurance, traffic safety, urban SS, SS talent training, SS knowledge popularization, SS media undertakings, etc., in the future. These important application research fields may also form a big disciplinary group in the future, and are shown in Table 2.

Research directions of the Intelligent Domain Risk Governance

From the analysis of Fig. 5, the basic disciplinary directions of the Intelligent Domain Risk Governance can be inferred, and are described in the following sections.

Intelligent social risk management

This is a research field that aims to systematically identify, analyze, predict, and respond to complex risks arising from the development of intelligent society through the deep integration of intelligent technology and social sciences, emphasizing the flexibility of risk response mechanisms, and balancing innovation and security. Its research topics include: focusing on new risks formed by the interweaving of technology, society, economy, ethics and other factors in an intelligent society, specifically, algorithm bias, data privacy disclosure, AI out of control, cyber security threats, technology unemployment, digital divide expansion, social trust crisis, online public opinion out of control, public infrastructure vulnerability caused by technology dependence, global supply chain disruption risk, etc. The core method is to use big data real-time monitoring, machine learning prediction model, traceability and other technologies to improve the ability of risk identification and dynamic assessment. The society is regarded as a complex adaptive system composed of people, technology and institutions, and the emergence, chain effect and cascade diffusion mechanism of risks are studied. In risk assessment and response, the efficient calculation of AI and human value judgment are combined. A risk profile through multi-source heterogeneous data, use digital twins, social computing simulation and other technologies is built, and the risk evolution path is deduced.

Table 2. Some important application research directions of the SS Intelligence Engineering and examples[38–44].

Directions	Examples
Applying AI to national security	Strategic threat deduction, multinational game simulation system, dynamic evolution model of geopolitical conflict, multi-source heterogeneous data threat intelligence extraction, digital frontier defense, cross-border data flow supervision intelligent system, cognitive warfare confrontation engineering, forged content traceability and countermeasures system, etc. More new SS disciplines that may be formed in the future include national digital security engineering, strategic intelligence deduction science, data frontier defense engineering, etc.
Applying AI to military security	Ethical constraints of weapon systems, dynamic regulation mechanism of AI independent decision-making, intelligent deception of battlefield environment, virtual technology of confrontation network, resilience enhancement of bee colony system, anti-jamming self-healing communication network of UAV cluster, confrontation of neural electronic warfare, encryption protection system of brain computer interface, etc.
Applying AI to social SS	Time-space prediction of crime hotspots, dynamic generation technology of crime probability, pre intervention technology of group event evolution, intelligent group behavior dredging system, hidden network ecological governance, anonymous network penetration analysis technology, public opinion immune engineering, false information group immune strategy, etc.
Applying AI to natural disaster prevention and reduction	Multi hazard coupling early warning, large model of earth disaster prediction, AI adaptive emergency, UAV disaster relief materials autonomous distribution system, urban flood disaster simulation and plan optimization, AI ecological resilience assessment, intelligent diagnosis of ecosystem vulnerability, etc.
Applying AI to public health emergencies	Pathogen evolution prediction, virus mutation path deduction, medical resource circuit breaker warning, medical run risk prediction, vaccine social dynamics, multi-agent population immune process simulation, non-contact epidemiology, spatial voiceprint identification, and intelligent tracking of connectors, etc.
Applying AI to accident prevention	Entropy increase monitoring of industrial systems, failure precursor recognition of multimodal data, intelligent prediction of human error, biological indicators tracking operators' safety status, chain accident propagation path cutting algorithm, accident investigation enhanced virtual reality, on-site debris intelligent reconstruction and cause tracing, etc.
Applying AI to SS supervision	Intelligent processing of regulatory big data, intelligent semantic analysis of regulatory provisions, off-site intelligent inspection, remote law enforcement system based on the Internet of Things, security regulatory deduction, digital testing of policy impact, enterprise risk profiling, enterprise security credit evaluation model, etc.
Applying AI to enterprise safety production	AI safety management system, risk self sensing system of the whole production process, gene analysis of violation behavior, intelligent analysis driving violation evolution, supply chain resilience assessment, risk conduction blocking algorithm, safety investment optimization, risk income balance model, etc.
Applying AI to emergency rescue	Multi-mode disaster perception, development of space integrated reconnaissance robot, digital twin of rescue force, AI collaborative command platform of on-site rescue force, adaptive rescue path, real-time calculation of three-dimensional space traffic capacity, intelligent triage of the wounded, and vital signs priority treatment decision-making system, etc.
Applying AI to safety technology R&D	Safety experiments, intelligent search algorithm for safety hazards, cross agency safety instrument and equipment, safety technology ethics evaluation, double-edged sword effect prediction model of innovative technology, etc.
Applying AI to traffic safety	traffic dynamic risk field theory, risk propagation modeling under the vehicle road collaborative environment, driving brain cognitive modeling, trust transfer mechanism between driver status and automatic driving system, intelligent repair of road damage of self sensing materials, intelligent generation of emergency channels, intelligent restructuring algorithm of traffic flow under major accidents, etc.
Applying AI to urban SS	Urban safety ecological intelligent monitoring, urban lifeline system health diagnosis based on the IoT, three-dimensional risk governance, resilient urban evolution, urban planning disaster resistance iterative optimization, community safety immunity, risk adaptive grid management system, etc.
Applying AI to safety personnel training	Safety knowledge evolution modeling, virtual safety tutor, intelligent planning of safety talents, emergency brain training, crisis decision-making ability training, etc.
Applying AI to safety science education	AI risk awareness remodeling, public safety mental model building, immersive education theater R&D, accident scenario AI experience system, intelligent safety short video production platform, intelligent quantification of safety education effect, etc.
Applying AI to SS media	Safety news event presentation, intelligent interference technology of negative information transmission path, media safety survival early warning, safety media health intelligent diagnosis, etc.

Intelligent technology system risk management

This is a research field focusing on the identification, assessment, prevention, control, and governance of potential risks in the whole life cycle of intelligent technology systems. Its core is to ensure the safety, reliability, and controllability of technology systems through systematic methods, while balancing technological innovation and social responsibility. Specific research topics include: algorithm deviation, model vulnerability, code vulnerability, training data pollution, privacy leakage, data heterogeneity, other data risks, and the impact of technical defects on social security and their governance, unexpected behavior of the independent decision-making system, trust crisis in human-computer cooperation, and other systems out of control, external risks such as sensor failure, external data input distortion, malicious attacks, ethical conflicts, system security engineering theory, complex security system analysis, dynamic risk assessment model, security ergonomics, full life cycle management, uncertainty modeling, cross domain risk transmission, technical, and ethical contradictions, AI security standardization, etc.

Management of intelligent technology applied in SS

This is a research field focusing on SS prevention and control, risk governance, and compliance implementation of intelligent technology in practical application scenarios. Its core goal is to ensure that the development, deployment, and use of intelligent technology comply with safety norms, ethical standards, and social needs through a systematic approach, and avoid physical injury, social disorder, and trust crisis caused by technology abuse or failure. Specific topics include: AI algorithm failure, data pollution, system interaction vulnerabilities, environmental complexity, human-computer conflict, compliance and other risks, scenario-based security modeling, security utility trade-off theory, resilient system design, compliance framework, extreme scenario simulation, real-time monitoring and feedback, multi-party collaborative governance, etc. The main difficulties are the large difference in security requirements of scenarios, real-time requirements, hidden risk transmission, and vague responsibility definition.

Global intelligent collaborative security governance

This is a comprehensive research direction on how to respond to the super sovereign risks and common challenges of mankind caused by intelligent technology through transnational, cross domain and cross-cultural cooperation mechanisms in the context of the deep integration of globalization and intelligence. Its core goal is to build an inclusive, resilient, and sustainable global intelligent security governance system, and achieve technology dividend and risk sharing. Specific research topics include: complex

global risks that transcend the boundaries of a single country or in industry in the intelligent era, technological hegemony risks, survivability risks, global military conflicts, threats to the survival of human civilization, systematic imbalance risks, marginalization of some nations, cultural homogenization and the demise of cultural diversity, global public governance theory, multi center collaborative governance model, civilization compatibility security concept, technology power balance theory, digital civilization governance community, global risk mapping and early warning, hierarchical governance architecture, soft and hard rule layer, dynamic adaptation mechanism, etc. The main difficulties include the reluctance of some nations to sacrifice technological sovereignty for global interests, the dominance of civilization conflicts, the paradox between free riding and collective action, the solidification of technological generational differences, the contradiction between efficiency and fairness, the contradiction between openness and security, and the contradiction between unity and diversity.

AI socio-technical risk prediction

This is a research field focusing on prospective identification, dynamic modeling, and quantitative assessment of future potential risks in the interaction between AI technology and social systems. Its core goal is to predict the impact of AI technology development on social structure, economic operation, ethical order, and ecological environment, and provide decision-making basis for future risk early warning and early active intervention. Specific research topics include: complex risks emerging from the two-way interaction between AI technology, and the social system, technology driven social risks, social feedback technical risks, ethical constraint backbiting, cultural adaptation crisis, resource crowding effect, complex adaptive system theory, causal inference and counterfactual prediction, multi-agent social simulation, resilience threshold theory, cross modal risk perception, prediction model architecture, extreme event modeling, uncertainty quantification, etc. The main difficulties include data heterogeneity, implicit researcher values in the prediction model, risk early warning that may change social behavior, and the correlation modeling between micro technical decisions and macro social effects.

After continuous enrichment and consolidation, the above research directions will hopefully become a group of new sub-disciplines of the Intelligence Domain Risk Governance.

Conclusions

The development and applications of AI have always been accompanied by SS problems. At the same time, AI also plays an important role in enabling various SS fields, and shows broad application prospects. In the AI outbreak period, its SS problems are becoming increasingly prominent. In the future, AI development and its universal applications will generate more new SS problems. The intersection of AI and SS represents a significant gap in the literature, as the symbiotic development of these fields for disciplinary construction has yet to be investigated.

The problems related to AI and SS can be divided into three-level and three-category problems: the first level is the SS problems of the AI system itself; the second level is the SS problems of the applications of AI to enable various fields; and the third level is the new SS problems derived from AI R&D and its extensive applications. The three levels and three categories of SS problems are interrelated and mutually coordinated, which need to complement each other to prevent AI from becoming out of control in the future.

According to the three levels and three categories of AI and SS, three new AI SS disciplines should be constructed, that is, the Intelligent System SS, the SS Intelligence Engineering, and the Intelligent Domain Risk Governance. By the given definition, connotation, research objects, knowledge structure systems, and interdisciplinary and difference analysis of the three new disciplines, it can be seen that these three new disciplines are very important and need to be put on the agenda and constructed urgently worldwide.

The general and special research directions and typical topics of the three new disciplines are summarized, and they are expected to form a new group of disciplines by the symbiosis of AI and SS science in the future. This investigation fills the theoretical gap in the construction of new disciplines of AI and SS, and has deep significance for the sustainable and safe development of AI, and the prevention and control of future AI risks.

Author contributions

The authors confirm their contributions to the paper as follows: studied and proposed the whole outline and content of the paper, studied the concepts and frameworks of the three new disciplines, and designed and wrote most of the content of the article: Wu C; participated in the construction of the three new disciplines, and wrote part of the outlook of AI SS: Wang B; participated in the document review and created the figures used in the article: Shi Z. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

This work was supported by the Education and Teaching Reform Research Project of Central South University (2024jy116), and Central South University Graduate Education and Teaching Reform Project (2025JGB046).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 24 December 2025; Revised 30 January 2026; Accepted 9 February 2026; Published online 23 March 2026

References

[1] Wu C. 2025. Research on the training mode of safety professionals for new quality productivity. *Safety & Security* 46:64–73 (in Chinese)

[2] Turing AM. 2007. Computing machinery and intelligence. In *Parsing the Turing Test*, eds. Epstein R, Roberts G, Beber G. Dordrecht, Netherlands: Springer. pp. 23–65 doi: 10.1007/978-1-4020-6710-5_3

[3] McCarthy J, Minsky ML, Rochester N, Shannon CE. 2006. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine* 27:12–14

[4] Sarker IH, Furhad MH, Nowrozy R. 2021. AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science* 2:173

[5] Johnson B. 2022. Metacognition for artificial intelligence system safety – an approach to safe and desired behavior. *Safety Science* 151:105743
[6] Lv ZF, Ma G, Sun XW, Wang LR, Shi LY, et al. 2019. Overview of concept, classification & study status of artificial intelligence security (I). *Smart Power* 47:32–42 (in Chinese)
[7] Wang HH, Li HM, Peng GH. 2024. Review of artificial intelligence security research. *Electronics Quality* 24:114–117 (in Chinese)
[8] Chen H, Wang FY, Zeng D. 2004. Intelligence and security informatics for homeland security: information, communication, and transportation. *IEEE Transactions on Intelligent Transportation Systems* 5:329–341
[9] Chen H, Wang FY. 2005. Guest editors' introduction: artificial intelligence for homeland security. *IEEE Intelligent Systems* 20:12–16
[10] Kharazishvili Y, Kwilinski A. 2022. Methodology for determining the limit values of national security indicators using artificial intelligence methods. *Virtual Economics* 5:7–26
[11] Sanclemente GL. 2022. Reliability: understanding cognitive human bias in artificial intelligence for national security and intelligence analysis. *Security Journal* 35:1328–1348
[12] Park AJ, Tsang HH, Sun M, Glässer U. 2012. An agent-based model and computational framework for counter-terrorism and public safety based on swarm intelligence. *Security Informatics* 1:23
[13] Vaseashta A. 2014. Advanced sciences convergence based methods for surveillance of emerging trends in science, technology, and intelligence. *Foresight* 16:17–36
[14] Drăgoicea M, Badr NG, Falcão e Cunha J, Oltean VE. 2018. From data to service intelligence: exploring public safety as a service. In *Exploring Service Science*, eds. Satzger G, Patrício L, Zaki M, Kühl N, Hottum P. Cham: Springer. pp. 344–357 doi: 10.1007/978-3-030-00713-3_26
[15] Radulov N. 2019. Artificial intelligence and security. Security 4.0. *Security & Future* 3:3–5
[16] Nasare R, Shende A, Aparajit R, Kadukar S, Khachane P, et al. 2020. Women security safety system using artificial intelligence. *International Journal for Research in Applied Science and Engineering Technology* 8:579–590
[17] Mahor V, Rawat R, Kumar A, Garg B, Pachlasiya K. 2023. IoT and artificial intelligence techniques for public safety and security. In *Smart urban computing applications*, eds. Jabbar MA, Tiwari S, Ortiz-Rodriguez F. 1st Edition. Denmark: River Publishers. pp. 111–126 doi: 10.1201/9781003373247
[18] Srivastava S, Bisht A, Narayan N. 2017. Safety and security in smart cities using artificial intelligence—a review. *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, India, 12–13 January, 2017*. USA: IEEE. pp. 130–133 doi: 10.1109/CONFLUENCE.2017.7943136
[19] Tokody D, Albini A, Ady L, Raynai Z, Pongrácz F. 2018. Safety and security through the design of autonomous intelligent vehicle systems and intelligent infrastructure in the smart city. *Interdisciplinary Description of Complex Systems* 16:384–396
[20] Wang B. 2021. Safety intelligence as an essential perspective for safety management in the era of Safety 4.0: from a theoretical to a practical framework. *Process Safety and Environmental Protection,* 148:189–199
[21] Wang B, Wang Y, Yan F, Zhao W. 2022. Safety intelligence toward safety management in a big-data environment: a general model and its application in urban safety management. *Safety Science* 154:105840
[22] Yampolskiy R, Fox J. 2013. Safety engineering for artificial general intelligence. *Topoi* 32:217–226
[23] Wang B, Wu C. 2019. Demystifying safety-related intelligence in safety management: some key questions answered from a theoretical perspective. *Safety Science* 120:932–940
[24] Patriarca R, Di Gravio G, Cioponea R, Licu A. 2019. Safety intelligence: incremental proactive risk management for holistic aviation safety performance. *Safety Science* 118:551–567
[25] Shefer O, Laktionov O, Pents V, Hlushko A, Kuchuk N. 2024. Practical principles of integrating artificial intelligence into the technology of regional security predicting. *Advanced Information Systems* 8:86–93
[26] Gwala RS. 2025. The use of blockchain technology and artificial intelligence in cryptocurrency and medical technology: a comprehensive review. In *Driving Socio-Economic Growth With AI and Blockchain*. USA: IGI Global. pp. 147–184 doi: 10.4018/979-8-3693-8664-4.ch007
[27] Wu C. 2022. Research on the basic theory of science of new disciplines. *Technology and Innovation Management* 43:342–350 (in Chinese)
[28] Wu C. 2024. Theoretical investigation on the collaborative development of the safety & security undertakings and the low-altitude economy to emerge new quality productive force. *Safety & Security* 45:52–61 (in Chinese)
[29] Wu C. 2021. Research on basic theory of safety complexity science: laying a foundation for new highland of safety science. *China Safety Science Journal* 31:7–17 (in Chinese)
[30] Zhao JW. 2025. Model classification and system connection of artificial intelligence science and technology ethics review system. *Contemporary Law Review* 39:84–96 (in Chinese)
[31] Song BZ, Qin RB. 2023. The legal regulation on the application risks of generative artificial intelligence. *Journal of Shanghai University of Political Science and Law (The Rule of Law Forum)* 38:108–121 (in Chinese)
[32] Farahmand F. 2023. A system engineering approach to AI security and safety. *Computer* 56:118–122
[33] Yang X, Shu L, Liu Y, Hancke GP, Ferrag MA, et al. 2022. Physical security and safety of IoT equipment: a survey of recent advances and opportunities. *IEEE Transactions on Industrial Informatics* 18:4319–4330
[34] Wu C. 2025. Research on fundamental theory of macrosecurisafetyology. *China Safety Science Journal* 35:1–13 (in Chinese)
[35] Wu C, Huang X, Wang B. 2024. Glimpse of safety science development in China: a review of safety fundamental research and construction of six new postgraduate courses for safety majors by safety & security theory innovation and promotion Center of Central South University. *Safety Science* 169:106323
[36] Wu C. 2024. Review of fundamental theories and research prospects of emergency management science from the disciplinary perspective. *Emergency Management Science and Technology* 4:e024
[37] Wu C, Wang B. 2023. Theory of creating new disciplines of safety and security (SS) science and essentials of 40 practical examples. *Emergency Management Science and Technology* 3:1–12
[38] Wang B, Wu C. 2018. Safety forecasting science: a branch of safety science in urgent need to be established. *Science and Technology Management Research* 38:258–266 (in Chinese)
[39] Huang X, Wang B, Wu, C. 2022. Realizing smart safety management in the era of safety 4.0: a new method towards sustainable safety. *Sustainability* 14:13915
[40] Zhao S, Yang H, Kareck LT, Khan F, Wang Q. 2026. Data-driven fault detection and diagnosis in industrial process systems: a systematic review and perspective. *Reliability Engineering & System Safety* 270:112159
[41] Chen S, Chen A. 2025. E-M-A-M emergency management methodology and application. *Emergency Management Science and Technology* 5:e006
[42] Zheng X, Ren J, Tong X. 2025. Towards a systematic framework for the safe development of cities: resilience, intelligence, and sustainability perspectives. *Emergency Management Science and Technology* 5:e020
[43] Wu C, Wang B, Huang L. 2023. *Principles of Safety & Security Science*. 2nd Edition. Beijing: China Machine Press. pp. 1–30 www.cmpedu.com/books/book/5606939.htm
[44] Yan F, Li X, Wang B, Xie Y, Wu C. 2023. Exploring safety capacity from a risk and safety information integration perspective: connotation, dimension mining and dimensionality reduction. *Heliyon* 9:e21728