中国移动
China Mobile

AI+智慧城市
安全解决方案白皮书

中国移动通信集团有限公司
2024年9月

Table of contents

Issuer: China Mobile Communications Group Co., Ltd.

Prepared by: Information Security Management and Operation Center of China Mobile Communications Group Co., Ltd.

China Mobile Xiong'an Information Communication Technology Co., Ltd.

China Mobile (Shanghai) Information and Communication Technology Co., Ltd.

China Information Security Magazine

Beijing Venusstar Information Security Technology Co., Ltd.

Aspire Digital Technology (Shenzhen) Co., Ltd.

Chengdu Thinking Century Technology Co., Ltd.

Shanghai Jiaweisi Information Technology Co., Ltd.

Contributors: Wang Yun, Yuan Jie, Feng Guohua, Zhang Feng, Jiang Weiqiang, Cao Xuefeng, Huang Jing, Sun Haitao, Li Ziye, Wang Guangtao, Lu Xiaohu, An Baoyu, Qiu Qin, Dong Hang, Yuan Sheng, Wei Hua, Yao Fei, Yan Shao Wei, Yu Chaoxiang, Guo Zhongyuan, Gu Tiantian, Zhao Wei, Tang Shuanglin, Zhou Tao, Li Gang, Liu Le, Chen Wenbo, Huang Qin, Zhang Mingzhu, Shu Shouheng, Zhong Zhicheng, Liu Chao, Li Zhuo, He Shengwen, Zhong Li, Xi Shangcai, Yin Xianyang, Li Juntian, Zhou Tianxiang, Zeng Li, Li Zhaolei, Zhao Guangyi, Guang Guoqing

1. Introduction

Smart cities are an important tool for promoting the modernization of urban governance systems and governance capabilities. During the "Fourth Five-Year Plan" period, the country made special plans for the development of digital economy and smart cities. The huge value and significance of data in important application scenarios of the digital economy are emphasized and highlighted.

At present, smart cities are in the initial stage in the field of artificial intelligence technology. Smart cities face security risks at the network, data, and application levels, as well as artificial intelligence (AI) in smart cities. Intelligence itself introduces new security risks. Smart city AI security is related to individuals, organizations, and society. It will affect the public interest and even the national interest and cannot be neglected. The release of the Interim Measures for the Administration of Generative Artificial Intelligence Services by the Cyberspace Administration of China in 2018 has led to the The work has an executable management method, which means that the relevant technologies and businesses in the smart city have a systematic A systematic and sustainable development environment.

In order to help smart city operators such as governments or organizations at all levels deal with the above many artificial intelligence Security challenges and problems to achieve long-term healthy development of the digital economy. Urban AI artificial intelligence security needs, proposes AI + smart city security system architecture, through smart city AI risk prevention and AI empowering smart city security are two solutions to solve risks at all levels of the city. Taking Jiangxi Province and Zhongshan Smart City as examples, the implementation of their AI security solutions is explained Looking ahead to future urban development, the white paper proposes recommendations for the safe development of AI in smart cities. We hope to contribute to further promoting, popularizing and improving the concepts, methods, systems and applications of AI+ smart cities. strength.

2. Overview of AI+Smart City Security Background

2.1. Current status of AI+ smart city development

The country's 14th Five-Year Plan clearly states: "Strengthen the construction of digital society and digital government, and enhance

The digitalization and intelligence level of urban services and social governance has been improved. Against this background, local governments have actively promoted

Promote smart city construction, list it as an important development strategy, increase investment, government support and

Guidance plays a key role in the development of smart city construction. At present, smart city construction is in

In the rapid development stage, relying on the support of information technology and communication technology, artificial intelligence, big data,

The application of emerging technologies such as cloud computing provides more possibilities for smart cities.

According to the data of "China Smart City Market Forecast, 2023-2027", in 2023, China's smart city ICT

The market investment scale exceeds 870 billion yuan and is expected to exceed 1.1 trillion yuan by 2027.

The 2024 Government Work Report proposes to accelerate the development of new quality productivity and carry out "AI+"

Action. Artificial intelligence based on big models, big computing power, and big data is gradually becoming a digital tool for urban governance.

It has brought new momentum and new paradigms of digitalization, scientificization and advancement. It has not only brought efficient and intelligent urban management

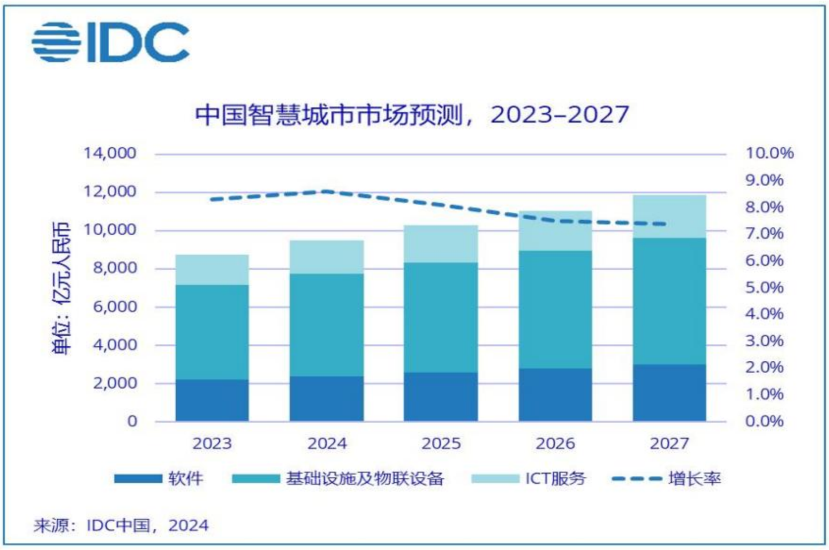
At present, governments at all levels in China

Under the guidance of national policies, we will actively explore new models of urban governance and integrate AI capabilities into smart city construction.

Video cloud human body detection, fire monitoring, smart operation and maintenance, intelligent question-answering robot, smart transportation and other smart

The construction of urban application scenarios is constantly improving, promoting the intelligentization of service scenarios such as urban governance and people's livelihood services

Chemical improvement.



来源：IDC中国，2024

Figure 1 China's smart city market forecast

The communications industry actively supports the country's promotion of AI+ smart city strategy. China Mobile, as a global leader

Communications and information service providers have clear strategies for the development of artificial intelligence and smart cities

planning and goals.

In 2023, China Mobile released the "China Mobile New Smart City White Paper", which focuses on

The key point is to build a new type of smart city through technologies such as artificial intelligence and big data.

The digital development strategy of the Party and the country is based on the deployment of a strong cyber power, a digital China, and a smart society.

The “Power Building” development plan was implemented, and smart cities were clearly defined as the key to my country’s urbanization development and urban realization.

The company focuses on building an open model of smart middle platform and integrating its capabilities into the strategic position of sustainable development solutions.

In smart cities, we will continue to improve basic communications, artificial intelligence, blockchain, and security based on business scenarios.

Full certification and other high-quality technical capabilities.

The 2024 China Mobile Artificial Intelligence Security White Paper points out that China Mobile has always regarded artificial intelligence as

It can serve as the company's strategic development direction, give full play to the advantages of operators, and create comprehensive artificial intelligence products

system, giving full play to China Mobile's resource and capability advantages in network, computing, and operation, and forming a comprehensive

Gaizhi computing power, high-quality data sets, artificial intelligence platform, algorithm capabilities in various fields, large models, intelligence

The full stack of new intelligent service capabilities of digital applications has reached the advanced level in the industry. China Mobile will accelerate the Integrate intelligence into smart cities, promote the development of "AI+" industries, and cultivate new quality productivity.

In the "State Council's Notice on Further Optimizing Government Services, Improving Administrative Efficiency and Promoting "Efficiently Doing One Thing" In the Guiding Opinions on "Things to Do", it is clearly required to explore and apply technologies such as natural language big models to improve online Intelligent customer service's intention recognition and accurate answering capabilities optimize intelligent question and answer, intelligent search, and intelligent guidance and other services to better guide enterprises and the public to do business efficiently and conveniently. Rich experience in building smart cities, based on a universal big model, integrating government data Fine-tuning, introducing the government affairs constraint model to restrict the output, creating an industry for the government affairs field Big Model - Jiutian Haisuan Government Affairs Big Model. With the government affairs big model as the technical foundation, we build a "platform + computing The "Law + Application" system has achieved an overall improvement and leapfrog development in the government affairs of Heilongjiang Province. One-stop service, one-network management, and one-network collaboration are three typical application scenarios to create intelligent customer service and Longzhengzhi search, digital human, official document writing and other specific applications, providing intelligent processing optimization, intelligent content content generation and other services to help the Heilongjiang Provincial Government perform its duties more efficiently and effectively improve the satisfaction of the people. Degree and trust.

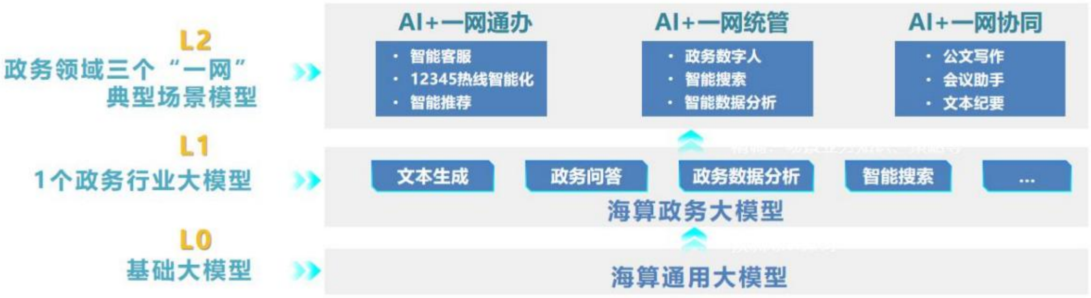


Figure 2 Application and practice of Jiutian Haisuan government affairs big model

2.2. Current status of AI+ smart city security development

The new smart city is a city that promotes the modernization of urban governance systems and governance capabilities and improves the well-being of urban residents. New concepts and new paths for happiness and satisfaction are also important for building a cyber power and developing the digital economy.

With the continuous development of AI technology and its wide application in the smart city field, people enjoy the benefits of technology. While reaping the benefits of technology, the company is also facing increasingly prominent, complex and ever-changing cybersecurity risks, including large-scale privacy Security incidents such as data leaks and cyber attacks on critical information infrastructure are common.

At present, there are two current situations in the development of AI+ smart city security. On the one hand, smart cities introduce large AI technology and various convenient technologies are profoundly transforming the operation and management of cities. With the improvement of convenience and efficiency, most city managers do not pay attention to the safety of AI itself: how to deal with it? Managing data security, privacy protection, and algorithm transparency are becoming important issues we face today. On the other hand, from the perspective of network attack protection in smart cities, facing new scenarios, new features, New demands, deep vulnerabilities and unknown threats are increasing, especially as the network security environment becomes more complex. The integration, interweaving and automation of information technology require the combination of AI to strengthen intelligent monitoring, early warning and proactive security protection. Active defense with linkage, front-line movement and coordinated linkage is particularly important.

Today, the security models at home and abroad are developing rapidly. The international market has seen the emergence of Cisco security artificial intelligence ÿÿÿÿElastic AI Assistant for SecurityÿGoogle Cloud Security AI Workbenchÿ Microsoft Security Copilot and other products, domestic manufacturers include 360, Anheng, Jinjing Yunhua, Huawei, Green Alliance, Qi'anxin, Venustech, Sangfor, and Topsec all proposed large or small security models. However, there are almost no training sets for the security features of smart cities. Urban safety is still in its infancy.

360 Security Group will release the Security Big Model 3.0 at the end of the first quarter of 2024, including language, planning, The five hubs of judgment, morality and memory, it is reported that the use of 360 security big model can achieve MTTR reduction Half, and the average work efficiency per person increased by 30%.

Huawei released the L4 AI security agent (cybersecurity highly autonomous) at the HSA2024 conference defense). Huawei has full-stack large model capabilities, including AI chips (Ascend series), CANN design computing architecture, MindSpore deep learning framework, MindStudio development tools to intelligent computing training network and

and Pangu Universal Large Model, which currently automatically handles more than 90% of security incidents, greatly reducing manual work

workload and improve the efficiency of network security operation and maintenance.

China Mobile and Venustech jointly released the Jiutian Taihe Security Model at the 2024 MWC conference.

The model relies on the powerful computing power and extensive data processing advantages of China Mobile's Jiutian Base Model.

Deeply integrate the massive data resources unique to the security industry of Venustech, including but not limited to threat intelligence,

Vulnerability database, professional security knowledge and the latest security research results.

Based on the usage scenarios, multiple small models are trained, such as medical data recognition and government data desensitization.

2.2.1. The urgency of preventing risks from **AI** technology in smart cities

With the widespread application of artificial intelligence (AI) technology in smart cities, not only the urban operation form

Great changes have taken place, and people's lifestyles have become more convenient.

The uniqueness of smart city resources and the potential security risks that may be brought about by complex and unpredictable AI technologies

Bad power is also unique.

Once AI decision-making in smart city services goes wrong, it may bring extremely bad social impact.

Affect the credibility of smart cities, and may even affect various industries, leading to social security, legal liability,

Moral dilemmas or negative impacts of public opinion may occur. For example, in park security management and disaster emergency management,

Once a problem occurs, a serious safety accident will occur.

It is imperative to prevent risks of AI technology in smart cities. We need to formulate effective strategies and measures.

Ensure that artificial intelligence plays a positive role while minimizing its potential risks. Establish a sound data security

Mechanisms and privacy protection policies should be established to ensure that personal data is used and protected appropriately.

Supervision and review of algorithms to promote algorithm transparency and fairness.

The development of smart cities insists on combining technological innovation with risk management, promoting the development of artificial intelligence technology

Sustainable development will improve urban governance and service levels. Only by fully understanding and effectively responding to AI technology can

Only by eliminating risks can smart cities achieve sustainable, intelligent and safe development.

2.2.2. Necessity of AI-enabled smart city security

In the context of smart cities, traditional network security defense technologies such as rule-based and authentication

Mechanisms such as firewalls, intrusion detection systems, and intrusion prevention systems play an important role.

However, these methods may not be able to cope with new threats and variant attacks.

Huge resource exposure, complex business logic, and diverse data transaction forms make security management

Faced with great challenges.

With the rapid development of network technology and data value, network security threats have become more intelligent and hidden.

The trend of anonymity and scale has brought greater challenges to the security defense of smart cities.

Security attacks enabled by artificial intelligence have become a common phenomenon, including but not limited to automatic vulnerability mining,

Intelligent malware generation, intelligent network destruction, etc.

Considering the characteristics of smart cities, AI technology is used to assist network security management to cope with the new situation.

By analyzing data streams with AI, it is possible to identify those that are difficult to detect with traditional methods.

Detect complex threats, thereby improving the accuracy and efficiency of threat detection.

Identify malicious code features to quickly and accurately detect abnormal behavior and malware, and suppress them at the source.

In addition, generating real-time threat reports and performing trend analysis can help prevent

Detect potential attack behaviors and improve the response speed and effectiveness of network security.

It improves the intelligence and adaptability of security defense and strengthens the ability to respond to security threats quickly and efficiently.

Through continuous innovation and optimization,

Combined with intelligent defense methods using artificial intelligence technology, smart cities can better protect network security and promote

Promote information sharing and intelligent applications to achieve the smooth advancement of urban digital transformation.

3. AI+Smart City Security Risks and Requirements

3.1. Risks of **AI** technology application in smart cities

At present, smart cities provide a variety of service capabilities, integrating various artificial intelligence technology research and development and application scenarios. With the rapid development of new technologies such as artificial intelligence, the security risks associated with them are becoming increasingly severe. The main risks involve smart city model algorithms, smart city data elements, smart city AI services, smart Urban AI platform and smart city AI operations, etc.

3.1.1. Risks of smart city model algorithms

(1) Risks of bias and discrimination in urban model algorithms

In the construction of smart cities, the risk of bias and discrimination in AI model algorithms is an issue that cannot be ignored. Because the training data may contain implicit biases, the model may be unable to make predictions or decisions. Intentionally favoring certain groups or opposing certain groups. For example, in the allocation of public services, traffic management, If AI models are influenced by biased data, they may

Therefore, when using AI technology, we need to consider how to avoid

Avoid and correct these potential biases and discriminations to achieve a truly fair and just smart city.

(2) Ethical and legal risks of urban model algorithms

In the implementation of smart cities, AI models will bring ethical and legal risks. These include AI The processing of personal information by technology may involve privacy protection and may lead to legal disputes if used improperly. At the same time, the bias and discrimination that may exist in the AI decision-making process also touches on the ethical level. For example, it may lead to unfair treatment of certain groups in enjoying public services. When using AI technology in the city, it is necessary not only to strictly abide by relevant laws and regulations, but also to consider its possible to achieve fairness, justice and transparency.

Smart city AI models also abuse user preference data, resulting in information gaps in city services.

The ethical risk of users being exploited by big data.

(3) Risk of unexplainability of urban model algorithms

AI model algorithms in smart cities generally face the risk of being unexplainable. This means that despite

AI models can generate effective predictions or decisions, but the complex computational processes within them are often difficult to understand.

This "black box" feature increases the risk of using AI models to a certain extent, because

Because it makes errors or biases difficult to detect and correct.

If the AI system we use makes a wrong decision and we cannot understand the logic behind the decision, then it is difficult to

To find the root cause of the problem and make improvements.

Since the big model of the smart city represents the authority of urban services, its decisions and outputs are directly

Affecting the public interest. If the big model is misleading, the resulting logic needs to be traced back and explained.

Therefore, for digital government, it is necessary to solve the problem of unexplainability of large models and ensure its decision-making process

Transparency and fairness are not only technical requirements, but also a reflection of social responsibility and public expectations.

(4) Risk of reverse tampering of urban model algorithms

Smart cities provide a large number of public interfaces and services to the society. Some malicious attackers may

Analyze and understand how AI models work, find loopholes in them, and obtain the internal logic of their operation to achieve

This adversarial attack can destroy the normal function of the AI model and has a great impact on smart transportation.

or public security monitoring, and may obtain the internal logic of the AI model and then

Reverse attack.

Similarly, attackers can use the public service interface of the smart city as an entry point to

Exploitation of vulnerabilities in public components or lack of awareness among personnel, illegally obtaining deployed AI model algorithms

detailed information, including parameters, structure, functions, etc., which may lead to infringement of intellectual property rights or commercial opportunities.

Risks of confidentiality leakage etc.

3.1.2. Risks of smart city data elements

(1) Risk of illegal collection of smart city data

In the operation of smart cities, various AI models need to collect and train a large amount of urban data.

Smart cities rely on a large amount of data to drive public services, traffic management, energy planning, and other aspects.

These data often involve national government affairs and urban management data. If they are not legally authorized,

It may constitute illegal or criminal behavior, and in serious cases may even affect national security.

Smart cities also require a large amount of personal information of citizens. With the introduction of the Personal Information Protection Law,

This puts forward higher requirements for data collection and processing. This means that when collecting and using personal information

When using the website, the user's consent should be obtained and the scope of use should be limited.

Any act of obtaining or using personal information will be severely punished by law.

The AI models involved in the operation must strictly comply with relevant laws and regulations to ensure that all collected and processed

All data is collected legally and in compliance with regulations to ensure national security and the personal information of citizens.

Rights and interests shall not be infringed.

(2) Risk of abnormal training data in smart cities

The AI model of smart cities relies on a large amount of training data. If there are anomalies in these training data,

This may cause the model’s prediction results to deviate. Abnormal situations include data containing noise, incorrect labeling,

The outliers are those that are inconsistent with the majority of the data.

Frequent data may affect the learning effect of the model, resulting in poor performance in practical applications such as traffic management and government affairs.

Misjudgment occurs in services and other aspects.

(3) Risk of data poisoning and pollution in smart cities

In smart city AI models, a large amount of training data comes from public data.

Malicious attackers may inject harmful or misleading information into the dataset, affecting the model training process.

For example, an attacker might modify public data such as water and electricity usage, or add specific

Data with incorrect labels can cause AI models to have bias or misunderstandings when learning.

It may seriously affect the performance and accuracy of the model.

(4) Risk of smart city training data leakage

On the one hand, there is a large amount of sensitive data in smart cities, including residents' personal information, geographic location,

location data, traffic flow data, etc. Once this data is illegally obtained or misused, it may have a negative impact on individuals.

This can cause serious violations of privacy and even lead to criminal acts such as national security data leakage.

There are a large number of external service interfaces and systems in Smart City, and the heterogeneous complexity of multiple systems greatly increases

Therefore, when AI training models use sensitive data, they should focus on data leakage.

risk.

3.1.3. Risks of Smart City **AI** Services

(1) Risk of deviation in urban service output content

Smart city AI services rely on accurate predictions and decisions from AI models.

Deviations in output content may affect the efficiency and fairness of the entire city's operations.

If the output of a question-answering robot is affected by data bias or erroneous learning, such as in the government

Deviations in Jing's answers will lead to a loss of credibility and reduced quality of social services.

(2) Risk of illegal content generated by urban services

Smart cities often use AI content generation services to provide services to society. AI-generated or synthesized content

Due to its randomness and uncontrollability, it may lead to illegal content, discrimination, bias, privacy violations, etc.

The emergence of many problems such as leakage and content infringement has a great impact on the safety of life and property of the public, national security,

Ideological and ethical security pose a threat. Especially in interactive questioning scenarios with large models, The prompt words entered by the user may also include those related to politics, pornography, terrorism, violence, gambling, drugs,

Risks of inducing crimes and malicious code and other illegal and irregular content. If the content security protection mechanism of the model

Imperfections may cause the model to produce output that is harmful, inappropriate, or contrary to public order and good morals.

Abnormal AI-generated content may cause the public information publishing system to mistakenly spread untrue or harmful information.

At the same time, if the content generated by smart city AI contains illegal or irregular content, it will also have an impact on the city's operations.

It will have an impact on operators and cause negative social opinion.

(3) Security risks of abnormal city service calls

Smart city AI service abnormal call security risks mainly involve service call failure, illegal access and

Frequent requests and other issues. In today's society, there are behaviors of using AI technology to attack the system.

AI can be used to batch generate data for large-scale smart city services, or AI can be used to identify patterns in the data.

Illegal access after bypassing identity authentication will pose a great risk to urban security and stability.

Traditional network security attacks are still effective against smart city AI services, including DDoS attacks, CC

Attacks, vulnerability scanning, Trojan horse implantation, etc. When network attacks affect city services on a large scale, it will lead to

If the business system is down, it will not be able to provide services normally, which will seriously affect urban traffic, public security,

Health emergency and other services have been suspended.

3.1.4. Risks of Smart City **AI** Platform

(1) Risk of malicious consumption of smart city computing power

The risk of malicious consumption of AI computing power in smart cities mainly comes from external malicious attackers and internal instability.

External attackers may cause AI server resources to be wasted by launching a large number of meaningless requests.

Large amounts of consumption will affect normal service provision. Inappropriate internal use, such as uncontrolled resource consumption,

Unreasonable task allocation may also waste computing resources on worthless tasks, affecting the

The processing efficiency of key tasks.

(2) Smart city supply chain security risks

Smart city AI technology platforms often rely on various system components developed at home and abroad.

Security threats may arise during the development, production, distribution and maintenance process.

At present, foreign technology is still the main technology, supplemented by domestic technology. Smart city AI technology may be affected by the international situation.

The security risks brought by factors such as the situation and technological blockade will ultimately affect the safety of the general public.

3.1.5. Risks of Smart City **AI** Operations

(1) Legal compliance risk

Smart cities must strictly abide by relevant laws and regulations in the process of data collection, storage, processing and transmission.

Laws and regulations, such as the Cybersecurity Law and the Data Security Law, ensure the legitimacy and security of data

However, with the rapid development of AI technology, its application in the field of smart cities is becoming more and more extensive.

This poses a severe challenge to existing laws and regulations. How to define the legal role of AI technology in smart city construction?

Legal responsibility has become an urgent issue to be resolved. In July 2023, the Cyberspace Administration of China and seven other departments jointly issued and

The implementation of the "Interim Measures for the Administration of Generative Artificial Intelligence Services" further regulates the development of AI technology.

To better utilize AI technology and avoid potential risks, we should adopt a more scientific and direct legal approach.

Carry out governance.

(2) Operational management risks

The construction of smart cities involves data sharing and collaboration across multiple departments, fields and levels.

With the development of big data, cloud computing, the Internet of Things and other technology stacks, the traditional operation model is prone to produce "chimneys"

Operation and maintenance requires high technical capabilities of operation and maintenance personnel. With the widespread application of AI technology, operations

The risks are becoming more prominent. First, smart cities are involved in the process of data collection, transmission, storage and processing, including

A large amount of personal privacy data of users, using AI technology to call and process data may result in data leakage

Second, the application of AI technology relies on a large number of algorithms. If the algorithms are wrong or out of control, Third, my country's smart cities still have unclear security operation models and reconstruction

Problems such as poor design and operation seriously affect the safe and efficient operation of smart cities.

Smart city security operations teams also face corresponding problems. The growing threat of cybersecurity is

This places an additional burden on the already stressed security professionals.

When it comes to cybersecurity issues, more experts specializing in AI and ML cybersecurity are needed.

The actual effectiveness of intelligent network security technology will be greatly improved by maintaining and adjusting it according to project requirements.

However, the number of qualified and well-trained professionals in this field worldwide is far from sufficient.

Meet current needs.

(3) Evaluating fuzzy risks

In the process of building a smart city, the operation evaluation of the smart city is a very important management

This will help us to continuously improve the construction and development of smart cities and promote our

The effective construction of smart cities requires grasping the basic direction of smart city construction. However, with the development of AI technology

Smart cities are widely used, and traditional smart city evaluation systems are not sufficient to support the application of new technologies.

New requirements.

3.2. Smart city security technology issues and **AI** empowerment needs

3.2.1. Smart city network security requirements

(1) Smart city intelligent network security requirements

For basic network security equipment, in the context of the development of artificial intelligence technology,

Unable to cope with the complex network environment in the next stage. The network environment of smart cities has sensor device data

The characteristics of large volume, multiple system types, and strong data mobility are not yet fully realized.

As the number of devices and systems connected to smart cities continues to increase, the attack surface also expands.

In addition, key infrastructure such as power grids involved in smart cities

Power, transportation, water supply, and communications systems are at risk of cyber attacks.

The traditional closed network security protection methods and threat based on features and rules are

It is difficult for threat detection technology to cover the complex network environment of the entire smart city.

(2) Demand for intelligent decision-making in smart cities

Big models are a widely used technology in the field of artificial intelligence.

Big model, among which the security big model related to the security field is a key technology in smart cities.

Traditional network security operation and maintenance methods require operation and maintenance personnel to have good technical capabilities to make

Timely and correct decision-making. In smart cities, the network environment is complex, there are many devices, and security operation and maintenance are difficult.

High, requires intelligent assistants that can support question-answering, intelligent question-answering models that can answer safe questions

Provide data support and decision-making suggestions for managers to help them make wise safety decisions

decision-making and improve the overall safety management level. By meeting the above needs, the intelligent question and answer model can significantly

Improve operational efficiency and protection capabilities in the field of smart city artificial intelligence security.

(3) Requirements for automated self-checking in smart cities

In the smart city network environment, in addition to the analysis and prediction of the network environment and real-time strategy

In addition, the security self-check of each system is also an important part. Conventional security tests such as penetration testing,

It is highly dependent on the human judgment and operation of technicians, and it is urgent to use artificial intelligence technology to assist testers.

Improve efficiency and threat identification rate. The intervention of artificial intelligence has advantages, especially in complex situations with multiple systems.

obvious.

3.2.2. Smart city data security requirements

(1) Demand for targeted processing of smart city data

In smart cities, a large amount of data will be generated and collected, including personal information, public service data, Improper use of these data may lead to serious consequences, such as personal privacy leakage,

Sensitive national and social data are exposed, and these data are maliciously used for advertising harassment, telecommunications fraud,

Social engineering attacks or business intelligence theft and other illegal activities. At the same time, improper data storage and use can also

This will lead to a series of risks. At present, smart city data lacks the necessary

The necessary data protection measures include the use of technologies such as data watermarking and data desensitization to protect smart city data.

Line labeling and limited desensitization make data traceable and prevent data generated when using artificial intelligence

Risk of poisoning and data leakage.

(2) Requirements for classification and grading of smart city data

In a smart city supported by artificial intelligence technology, data is an important asset, but the huge amount of data

This often results in a lack of organization and sorting of important or sensitive data, making it difficult to implement effective

Classification and grading protection measures. Using intelligent means to classify and grade massive data can not only improve

Efficiency, reducing manual assistance processes, and automated processes can make data sorting no longer cumbersome.

A variety of intelligent classifiers are used to automatically process labels in various industries in smart cities, providing a basis for artificial intelligence

Can provide a good data foundation.

(3) Smart city data compliance audit requirements

Smart city data audit compliance requirements are a comprehensive requirement to ensure that smart city construction

Ensure the security, compliance and validity of data in the facility, and comply with national and local regulations on data protection, privacy

According to relevant laws and regulations on data protection, a sound data management policy and process should be established to clearly define data access and

The authority and norms for access and use of data will be strengthened, and the monitoring and auditing of data will be strengthened. Intelligent audit systems and tools

Tools can facilitate the review of key data and strengthen the management and implementation of data auditing measures.

Ensure the security, compliance and effective use of data in smart city construction.

3.2.3. Security requirements for smart city applications

(1) Smart city content security requirements

In smart cities, applications cover traffic management, energy management, public safety, health,

Environmental monitoring and other aspects have greatly improved the efficiency of urban operations and the quality of life of residents.

Abnormalities in the content of Smart City software will reduce the efficiency of city services and may even affect public opinion.

On the credibility of smart cities.

(2) Smart city application monitoring requirements

Upper-layer applications and services support the operation of smart cities, providing multiple functional access points to facilitate

City service providers and users have faster access. Similarly, access points such as APIs give attackers

Provide attack channels. Artificial intelligence will provide more powerful technology for application monitoring and realize intelligent urban management.

Covering all APIs, dynamically monitoring API unauthorized attacks, privacy leaks, and denial of service attacks

It can also identify risks in interactive content, issue warnings in real time, and protect services and

The security of the application ensures continuous and reliable service.

(3) Smart city mimicry analysis needs

With the development of artificial intelligence technology, more human-like models and applications have been born.

While providing convenience to users, it also provides malicious attackers with new means of attack, allowing them to circumvent conventional defenses.

To combat such attacks, we also need the support of artificial intelligence technology.

Through machine learning, the behavior of malicious users and entities is analyzed to reduce the error rate of manual identification.

At the same time, technologies such as mimic honeypots can deceive attackers by simulating real services and disguising themselves to achieve

This type of AI security technology can fight against higher-level attacks.

Improve the overall defense capabilities of smart cities.

3.2.4. Smart city public security needs

Public security is related to national security and social stability.

With the increase of population, diversification of functions and continuous expansion of scale, the urban operation system is becoming increasingly complex and the security risks are increasing.

The traditional urban public security management has been unable to adapt to the requirements of the development of the times and cannot be effective.

To meet new challenges. The continuous development of AI technology has played an important role in long-term tracking, intelligent analysis, trend prediction and urban development.

The advantages of the city's precision management can help improve public safety risk situation awareness, prediction and warning,

Dynamic control and other capabilities. In addition, in high-precision recognition, real-time processing, traffic safety monitoring

Areas such as agriculture and rural areas also urgently need the application of AI technology to help improve the modernization of urban governance capabilities.

4. China Mobile AI+ Smart City Security System Architecture

Based on the national standard smart city system architecture and the China Mobile Artificial Intelligence Security White Paper,

Under the guidance of China Mobile, AI+ Smart City Security System Architecture is aimed at smart city AI risk prevention and AI

The design is based on two parts to empower smart cities.

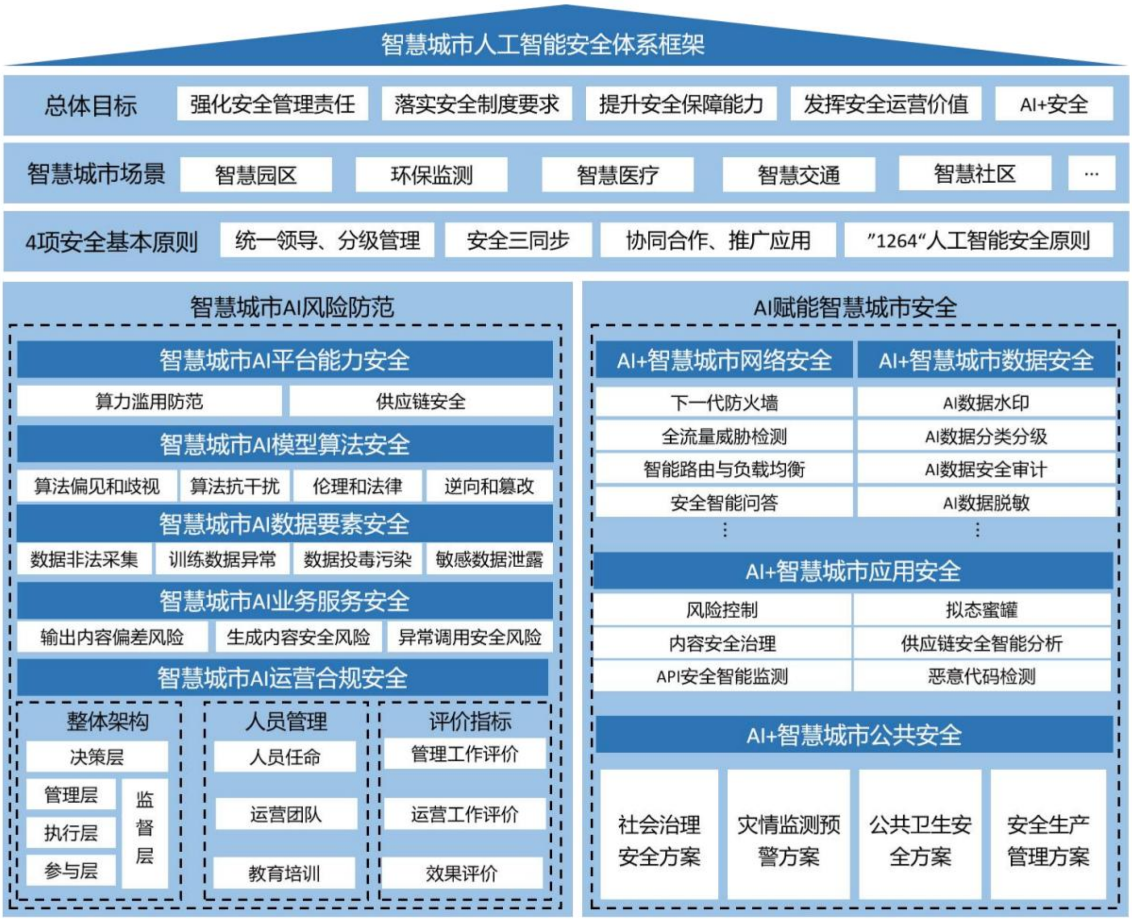


Figure 3 AI+ smart city security solution system framework

4.1. Overall objectives

AI+ smart city security is mainly reflected in smart city AI risk prevention, AI empowering smart city security,

First, make sure that the AI model is legal and compliant, the algorithm is fair and just, the data is secure and reliable, and the computing

The platform is manageable and controllable. Secondly, strengthen AI's empowerment of smart city security and apply intelligent means to smart

In the security protection work of smart cities, we will improve the level of network security. Ultimately, we will realize the overall goal of "making smart cities

Safer, and make cities safer and smarter".

4.2. Risk prevention **of AI** in smart cities

Smart City AI Risk Prevention: Comprehensively analyze the security risks of artificial intelligence technology applications and platforms.

The five aspects of the project are considered: model algorithm, data elements, business services, platform capabilities, and operational compliance.

Through the review and optimization of model algorithms, data privacy protection, business service monitoring,

Platform security enhancement and compliance management ensure that AI capabilities in smart cities are safe and controllable.

Such comprehensive risk prevention measures not only improve the security of smart cities, but also

Ensure the stability, reliability and sustainable development of AI technology, and promote the construction of smart cities towards a safer

And keep moving forward in a reliable direction.

4.3. **AI** empowers smart city security

AI empowers smart city security by conducting research on core artificial intelligence technologies in various scenarios of smart cities.

Applications in the field of security, such as big data models, can improve basic network security and data security governance.

The protection levels of management, content security governance, and business application security.

AI empowers smart city security by conducting research on core AI technologies and applying them to

In various scenarios of smart cities, the city's security level can be improved.

Core technologies can strengthen basic network security, data security governance, content security governance, and business applications

Use safety and other protective measures.

In terms of basic network security, AI technology can achieve real-time traffic monitoring and intrusion detection, identify

In terms of data security governance, AI can be used to

Encryption, access control and security auditing ensure the security of data transmission and storage.

Governance, AI can detect malware, filter harmful content, and ensure a clean and healthy network environment.

In terms of business application security, AI technology can reduce security risks through identity authentication, access control and other means.

Reduce all vulnerabilities and data leakage risks to ensure the stable operation of business systems.

Through the comprehensive application of AI technology, smart cities can improve security protection levels and reduce network risks.

AI-enabled smart city security

The system not only improves the intelligence and adaptability of security prevention, but also provides a better

Providing efficient and reliable solutions to promote smart city construction towards a safer, smarter and more sustainable

direction of development.

4.4. Basic principles of smart city security

The security of smart cities will follow the following four principles:

(1) Unified leadership and hierarchical management: Smart city security follows the principle of "whoever is in charge is responsible; whoever is in charge is responsible".

The principle of "whoever operates is responsible; whoever accesses is responsible" is used to clarify the division of responsibilities and implement manual

The main responsibility for intelligent security.

(2) Three synchronizations of safety: In accordance with the institutional requirements of the Ministry of Industry and Information Technology on the three synchronizations of safety,

In the process of building and operating smart city AI security, it should comply with the principles of synchronous planning, synchronous construction,

Three principles of synchronous operation.

(3) Adhere to the "1264 AI Security Plan": China Mobile has made clear the AI security

The development principle of the field is to plan a work system framework, focus on two work directions, and implement

Six major AI security risk protection measures, clarifying four categories of AI-enabled network security work. Subsequent AI+

The safe development of smart cities will also follow this principle to build a multi-level, comprehensive security assurance system.

(4) Adhere to collaborative cooperation and promote technology application: Maintain an open and cooperative attitude and actively participate in the industry.

Research on AI safety standards in the industry, carry out extensive cooperation and resource sharing, and jointly face the AI+ strategic transformation process

We will overcome new challenges in the process and jointly reach new heights in AI+ smart city security.

5. Smart City AI Risk Prevention Solution

5.1. Smart City AI Model Algorithm Security

5.1.1. Maintaining fairness and transparency in the city model

In the design of model algorithms, it is necessary to focus on selecting features and parameters, and construct A test dataset containing various scenarios and sources was created to fully test and verify the model algorithm to ensure The decision results for all objects are consistent. It is necessary to focus on evaluating the fairness of the model, such as Fairness evaluation based on static test datasets.

In order to comply with laws and regulations on algorithm transparency, it is necessary to establish a transparent A regulatory algorithm management system should be established, and the mechanism of the algorithm should be made public as required. Smart cities can adopt Vision and language-assisted explanation, policy imitation, interpretable models, logical relationship extraction, and policy decomposition, etc. Technical means are used to explain the model's reasoning process, which can enhance the model's transparency.

5.1.2. Improving the interpretability of urban models

Improving the interpretability of smart city AI model algorithms requires the use of clear and transparent algorithms and recording Decision-making process and parameter setting, provide visual display results and explanation mechanism, establish model documentation and explanation Instructions, regularly review and update algorithms, and strengthen training for users and related personnel so that they Ability to understand and interpret model results and ensure that the decision-making process is traceable and explainable to maintain the legitimacy of the model. Rationality and fairness. Using visual and language-assisted explanations, strategy imitation, interpretable models, logical relationships Technical means such as system extraction and strategy decomposition are used to explain the reasoning process of the model to enhance the transparency of the model.

5.1.3. Ensure that the city model is legal and compliant

To ensure that the ethical and legal risks of AI model algorithms in smart cities are resolved, it is necessary to clarify the privacy of various types of urban data.

privacy protection policies, comply with data compliance requirements, establish a transparent algorithmic decision-making process, and conduct risk assessments.

According to the national "New Generation Artificial Intelligence Ethics Code" and "Science and Technology Ethics Review Measures (Trial)"

The 1999 Science and Technology Law stipulates that regular ethical review and inspection should be carried out to strengthen the prevention and control of scientific and technological ethical risks and promote responsible

Be innovative and effectively prevent potential risks. Increase service usage feedback communication for smart city users

A channel through which users can provide feedback on issues encountered while using the AI system.

Smart city managers should pass the "Internet Information Service Algorithm Preparation" of the State Internet Information Office.

The service model and service form used by the service are recorded in the "Record System". Ensure effective supervision and

Transparent use.

5.1.4. Add city model encryption and obfuscation

To ensure that the smart city model is irreversible, model encryption and obfuscation technology can be used to

The data is encrypted and obfuscated to protect the model from malicious cracking.

Desensitization and de-identification can reduce data relevance and ensure the anonymity of output results.

This prevents the model from being reversed.

5.2. Security of **AI** Data Elements in Smart Cities

5.2.1. Smart city data collection security

In order to ensure the security of data collection for smart city models, it is necessary to verify the credibility of the data source and ensure that the data is

Ensure that the data source does not contain any illegal or harmful information, bias, discrimination, or commercial

Secrets, etc., and mark the data source for traceability or obtain open source license agreements. At the same time, clear compliance

According to the data collection, use and storage rules, sensitive information is desensitized to ensure personal privacy

The necessary user notification form can be signed according to compliance requirements.

Smart cities need to determine the scope of users or programs that can access data related to AI systems.

Data security access control is implemented. To implement these controls, account numbers and

authentication methods such as password, fingerprint recognition, and face recognition, as well as role management, permission management, and access control.

At the same time, smart cities need to conduct data access control.

Auditing includes recording information such as access user information, access time, access content and access results.

5.2.2. Smart City Training Data Configuration

To ensure the rationality of the training data configuration of the smart city model, we must first clarify the model objectives and select

Secondly, appropriate data preprocessing, such as removing noise, filling

Fill missing values, perform feature engineering, etc. to improve model performance. Maintain the diversity and peace of the data set

Balance to avoid bias that may lead to inaccurate model training. In addition, the model should be evaluated and adjusted regularly.

Continuously optimize the parameter settings to ensure the rationality and effectiveness of the final model effect.

Track new research results and technological developments, and update and optimize data configuration solutions.

To enhance the security of data annotation, before annotating training data, it is necessary to clarify the purpose of annotation,

The annotation content, the qualifications of the annotation personnel, the annotation environment, and the type and level of the original data are to ensure

Traceability of the annotated content. During the annotation task, security audits and data classification can be performed.

Store and review the annotation process. When the annotation results are output, we check the format,

The level and content are verified, and security measures such as encrypted transmission are taken when data is delivered.

After the labeling task is completed, the data labeling situation needs to be manually checked.

If it is accurate, it will be re-marked.

5.2.3. Smart city data poisoning prevention

Ensuring that smart city model training data is not contaminated by poisoning mainly relies on effective data cleaning

and verification to ensure the reliability of data collection sources and prevent malicious activities by strengthening access control and authority management.

Insert outliers. Use machine learning or statistical methods to detect and remove outliers or significant deviations from the norm.

The distributed data should be regularly evaluated for model training results. If there is a decrease in prediction performance or model bias,

In case of abnormal situations such as large differences, the data needs to be checked immediately and necessary cleaning and filtering must be performed.

During the model development phase, smart cities can introduce adversarial examples by adding small perturbations or

Data enhancement is performed to improve the model's anti-attack ability. From the perspective of model structure, we combine multiple

The outputs of the models are fused so that even if some models fail to provide service, we can still

At the same time, smart cities can regularly update models and add new training

data to maintain the robustness of the model. When the model robustness decreases, we will promptly perform model

updates and optimizations to ensure the stable operation of the model.

5.2.4. Smart City Data Leakage Prevention

To prevent the leakage of smart city training data, security management and technical means can be used, including access control.

Access control, data classification and grading management, and encrypted storage and transmission of sensitive data to ensure that training data

Security in storage, transmission, processing and use. In the use stage of training data, smart

The city business system can record key operational behaviors.

Model training and testing in environments with insufficient security measures, desensitizing or de-identifying sensitive information

After the training and testing, the relevant data will be securely transferred or deleted.

5.3. Smart City **AI** Business Service Security

5.3.1. Monitoring of urban service content generation

Smart cities need to develop a content security management and monitoring mechanism that contains harmful information.

The mechanism is based on the relevant rules and regulations on the production and dissemination of AI content, and combines various illegal

The content is graded and classified to manage the possible harm caused by bad information.

In the user input stage, smart cities target those who input illegal and harmful information or use,

The model generates and disseminates illegal and negative information, and sets refusal to answer or further punishment measures.

In the model output stage, keywords or sensitive word libraries are set, and classification models and other methods are used to classify the output content.

At the same time, smart cities also need to set up criteria for judging abnormal answers and normal answers.

You can set up a standard question library to identify specific questions and call standard answers to reduce output.

We will not cooperate with content that is ideological, biased, or infringes on the rights of individuals or organizations.

Finally, by building a content security monitoring system, managers can monitor the expected output or

Monitor the disseminated text, pictures, audio, video and other content, and conduct

Filter and manually assist in reviewing and processing suspected bad information.

5.3.2. Identification of fake content in city services

Smart cities need to pay attention to the identification of fake content in their public opinion monitoring work. Deep fake is a

Common illegal and irregular behaviors on social media, which use artificial intelligence technology to tamper with and fabricate real

Images, videos and audios can mislead the public and have adverse effects.

According to industry regulatory requirements and business development needs, you can deploy

A detection system for synthetic content, which can generate common deep synthesis algorithms and artificial intelligence models

or synthesized content.

5.3.3. Smart city service call security

Smart city service call security mainly involves identity authentication, permission control and data transmission security.

Authentication ensures that only authorized users or services can access the system, preventing malicious attacks.

Control is performed after identity verification to determine which resources a user or service can access and prevent abuse.

Transmission security involves the encryption and integrity checking of all data transmitted over the network to prevent the data from being

At the same time, there needs to be an emergency response mechanism to detect any security incidents.

Detect, record, analyze, and respond to protect the security and stability of smart city services.

Smart cities also need to protect against cybersecurity attacks. External access to their AI systems, input

Input data and behavioral decisions are tested to detect security attacks on business systems in a timely manner.

Smart cities continuously monitor the operating status and security of AI systems and promptly warn any system

In addition, smart cities can deploy AI application services

The risk monitoring and security protection capabilities of calling interfaces, including asset monitoring, access control,

Abnormal behavior monitoring, blocking, API interface encryption, dynamic robot interception, weak password protection, security

Full audit, and API operation status monitoring, etc.

5.4. Smart City **AI** Platform Capability Security

5.4.1. Preventing abuse of computing power in smart cities

Deploy computing power security control measures to prevent computing power resources from being used in illegal application scenarios, such as

Activities such as network attacks or password cracking, which are initiated by malicious attackers using powerful computing power. These measures include

By analyzing the computing task type and combining the computing power threshold of computing power users, smart city computing tasks are

If the computing power threshold is exceeded, the system will limit the computing power usage or reject the computing power request.

And consider reducing the user's credit. In addition, the system also audits the operations of computing power users and monitors

And record any abnormal behavior to ensure the correct and safe use of computing resources.

5.4.2. Smart City Supply Chain Security

Before training the smart city AI model, check the versions and known vulnerabilities of the components used.

Domestic computing power, promote the adaptation of artificial intelligence systems to domestic chips, and increase the proportion of independently developed computing power equipment

In order to ensure the security of the supply chain, Smart City

The city industry line has established a set of risk management, supplier selection and management, product development procurement and installation

A complete supply chain security management strategy including full maintenance. This not only covers all aspects of the supply chain

Risk assessment and control also includes strict screening and management of suppliers, as well as product development and procurement

Through these measures, we can ensure the safety and stability of the supply chain.

Reduce risks caused by supply chain issues.

5.5. Compliance and Security of Smart City **AI** Operations

In the digital age, the means and frequency of cyber attacks are increasing, and the security threats faced by enterprises are increasing.

The traditional security operation center (SOC) relies on manual analysis and response, which is not only inefficient,

It is also prone to false positives and false negatives. The expansion of the attack surface and the increase in data volume have caused security operations personnel to be overwhelmed.

It is difficult to respond accurately in a short time.

and natural language processing technologies can quickly identify potential threats in massive amounts of data and provide

Effective solutions to improve safety operation efficiency, maximize the advantages of the system, and meet the country's new

Requirements for secure operations in smart cities.

5.5.1. Legal Compliance Construction

Smart city safety operation system is the cornerstone to ensure the smooth progress of safety operation.

The construction of systems, management systems, and compliance systems will help smart cities clarify security management standards and processes

and division of responsibilities, timely discover and respond to network security risks, improve the efficiency of security operations, and ensure

At the same time, with the widespread application of AI technology, smart

The safe operation of cities has brought great convenience, but it also brings a series of security risks and challenges.

The unpredictability of artificial intelligence, its ability to evolve itself, and the violation of privacy and data are issues that need to be addressed.

Smart cities pose a threat to security and credibility, so a sound AI+security industry management system is imperative.

It must be done.

5.5.1.1. Safety system construction

Smart city security operation management system is the basis for doing a good job in security operation. Develop safety operation management systems and strategies based on the actual needs and specific requirements of the country, and Continuously optimizing and updating.

At the same time, in order to cover all levels of security operations, a hierarchical security operations management system can be formulated System, including but not limited to safety strategy, safety system specifications, safety system processes, safety rules and Guidelines and other management systems at multiple levels.

Security strategy documents are mainly formulated based on smart city security operation goals, business needs, etc. Safety operation management policy, guiding the construction objectives, management scope, basic principles and Safety system and regulatory documents are mainly formulated to implement policies and guidelines. Standards and regulations should be established to ensure safe operation management, personnel management, education and training, monitoring and early warning, Emergency response, safety assessment, inspection and evaluation system specifications. Safety system process documents are mainly used for Clarify the processes and standardized operations of safety operations management, solidify safety actions, and implement them as a system. Generally includes security incident management, information backup management, security training and assessment, authority management, emergency response Safety rules and guidelines are mainly used to guide specific operations or operations. The traceable documents during the operation process generally include application forms, safety reports, safety records, event lists, Account password and other files.

5.5.1.2. AI system construction

Artificial intelligence technology has a wide range of applications, including various systems, algorithms and models. Security issues involve data privacy, system vulnerabilities, etc. By establishing a reasonable management system, It can identify and resolve potential risks in advance and ensure the security of artificial intelligence technology.

(1) Strengthening data privacy protection

Data is the core resource of artificial intelligence technology, and the leakage of data privacy often leads to serious consequences.

Therefore, it is very important to establish a sound data privacy protection system, including clarifying the data collection and use

Standardize the use of data encryption technology, strengthen the research and development and application of data encryption technology, etc.

(2) Improving the regulatory mechanism for AI systems

A strict regulatory mechanism needs to be established for the use and operation of artificial intelligence systems.

This includes standardizing the training and testing processes of artificial intelligence algorithms to ensure the stability and controllability of the system.

(3) Strengthen the discovery and repair of security vulnerabilities

Security vulnerabilities in artificial intelligence systems are the root cause of their security problems.

Strengthen the discovery and repair of security vulnerabilities. Establish a rapid response vulnerability repair mechanism, and

Eliminate the threat to Hong Kong at the same time.

5.5.2. Operation management and construction

In building a smart city security operation system, operation management can effectively connect with the security management system.

The smart city

An important part of the safety operation system.

5.5.2.1. Smart City Security Organization and Operational Structure

In the process of building smart cities, due to its wide range of services and multiple scenarios, it will involve urban management.

The construction and operation by multiple parties such as the Management Bureau, Big Data Bureau, and third-party development companies need to be carried out according to the decision-making level, management level,

The organizational structure of the three layers, namely, the executive layer, the participating layer, and the supervisory layer, is rationally designed to organize the safe operation of smart cities.

Organizational structure to ensure effective coordination and integration of resources from all parties, and to ensure communication and collaboration between all levels and departments;

Guide and promote the formulation and implementation of safety management systems to ensure the implementation and execution of safety measures.

Add an AI operation group based on the traditional operation organizational structure, and use AI technology to automate some

Data analysis, forecasting and other repetitive tasks greatly improve the operational efficiency of enterprises and help institutions

Reduce labor costs. In addition, by optimizing algorithms and models, reduce operating costs.

Full operations bring new development opportunities.

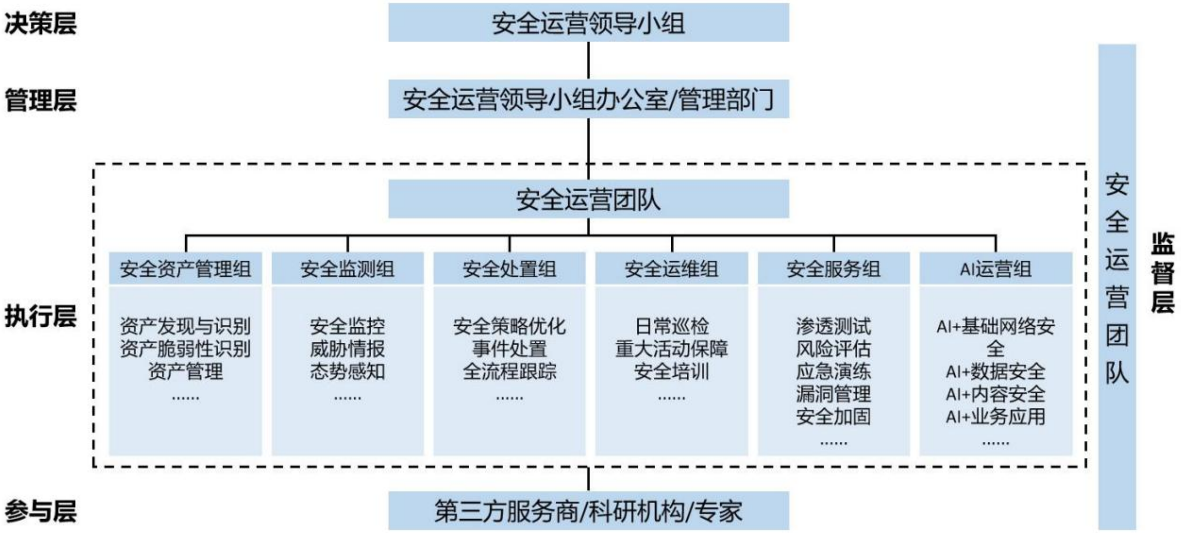


Figure 4 Security operations organizational structure

(1) Decision-making level

The decision-making layer is the top level of the smart city security operation organization and the decision-making body for security operations.

It is generally composed of the main leaders of the smart city and the highest person in charge of network security within the smart city.

The decision-making level is mainly responsible for coordinating and guiding the security operations in the construction of smart cities, formulating security operations strategies,

Responsible for the review of safety operation standards, management systems and other major matters.

(2) Management

The management level is the bridge between the decision-making level and the executive level. It is responsible for carrying out the tasks assigned by the decision-making level and

To guide and supervise the daily activities of the executive layer, and play a decisive role in the organizational structure of smart city security operations

The network security department and the information security department are usually responsible for the organization.

Based on the strategic guidelines given by the decision-making level, a detailed plan is formulated for the actual work of security operations, and measures are taken according to local conditions.

Formulate safety operation systems and standards that match safety operation plans; implement

Regarding the requirements for safe operation construction funds, implement the safe operation responsibilities and division of labor of each participant, and establish

Establish a safety operation management organization with clear responsibilities and powers; strengthen the construction of a safety operation team and strengthen personnel safety management.

Responsible for organizing personnel safety awareness, safety skills education and training, formulating safety assessment mechanisms, suppliers

Personnel security management mechanism, etc.; arrange, coordinate and supervise the network security work of each department and report to the higher level in a timely manner

Department reporting, etc.

(3) Execution layer

Considering the wide range of smart city services, the executive level is generally composed of business departments that provide specific security operations.

The executive level needs to conscientiously implement the safety operation

According to the requirements of the operation, formulate specific implementation plans based on the instructions of the superior departments and clarify various tasks of safe operation

The phased tasks, division of labor and time nodes should be ensured to ensure the orderly progress of work; daily safety operation and maintenance,

Emergency response, safety inspection, etc. At the same time, artificial intelligence technology (AI) is used to operate and

Management and optimization process, strictly abide by the safe operation procedures, and promptly discover the system specifications

Vulnerabilities and potential dangers help management make timely adjustments and improvements.

(4) Participation layer

The participation layer is the assistant of the management and execution layers. The participation layer is mainly composed of third-party service providers (such as

The participating layer is mainly responsible for coordinating

Assist the executive layer in undertaking the construction, implementation, maintenance and service of security operations; assist the management layer in undertaking

Development of safety operation standards or systems.

(5) Supervisory layer

The supervisory level must be independent from the decision-making level, management level, execution level, and participation level, and personnel cannot hold concurrent positions.

Ensure that its supervisory audit work is not affected by the other four layers, so that the organization can discover the security operations

The supervisory level mainly focuses on the systems,

Check, supervise and evaluate the implementation of strategies, normative documents, etc., and provide safety operations

Carry out supervision and implementation, and monitor and audit safety operation risks.

5.5.2.2. Smart City Security Organization Personnel Management

(1) Personnel appointment

People are the core element of an organization. The essence of personnel appointment is to assign the right people to the right positions.

The appointment of personnel is mainly based on the following aspects:

Types of roles: First, the person in charge is appointed. In the safety operation organization system, the first person in charge of safety is appointed.

Appoint a person to be responsible for security-related matters and bear relevant security responsibilities, usually the first person in charge of the department

The second is the Chief Information Officer (CSO), who is responsible for the company's overall information security strategy, formulation and

Implement information security policies and supervise all information security activities within the enterprise; third, network security work

Engineer, mainly responsible for monitoring the company's network, preventing and responding to network attacks, ensuring network stability and

security; fourth, security auditors, who assess the company's

Whether information security measures are effectively implemented and comply with relevant laws, regulations and standards.

(2) AI talent education

(1) The application of AI technology in the field of smart cities is becoming more and more extensive, which not only improves the intelligence of cities

level, and also brings more efficient and convenient management and service methods. However, to ensure that AI technology

To fully realize the potential of smart city compliant applications, we need to introduce professional AI technicians.

We also provide AI technology empowerment training to institutional managers and general operation and maintenance personnel.

The application of compliance in cities and give full play to its potential provide effective support for the sustainable development of smart cities.

We will cooperate with AI-related majors of well-known universities to jointly train talents.

Joint laboratories or special research groups, sending teachers, students and personnel to participate in them, and going deep into the teaching site

We will work closely with the frontline business to combine academic research, professional teaching and smart city business development needs, and provide more targeted

Cultivate future talents from the source in a targeted manner.

5.5.2.3. Smart City Security Operation Model

Smart city security operation solutions use AI big models to help enterprises solve the number of alarms in network security

There are many problems, such as too much noise, and the staff's ability is difficult to support efficient alarm and event analysis.

Monitor the abuse of AI technology in smart cities. The security operation model has natural language dialogue capabilities.

Intelligent functions such as detection capabilities can achieve 7×24 hours all-weather duty, improve efficiency and shorten

Response time to network security risks and threats. The application of large models carries 80% of security operations.

It greatly improves the efficiency of security operations, thus building a new paradigm for security operations.

The full operation model can realize "discovery and alarm - intelligent analysis - threat identification - blocking and isolation - impact investigation -

Strengthen the whole process of "reinforcement suggestions".

5.5.3. Evaluation indicators

(1) Security management indicators. These indicators mainly focus on whether the security operation and management measures of smart cities in various places are sufficient.

The evaluation mainly includes safety operation strategic planning, safety operation standard specifications, safety operation

There are five indicators: management organization, personnel safety management, and safety operation investment.

(2) Security operation indicators. They mainly focus on the security operation system of smart cities during operation.

The risk identification, safety monitoring and emergency response capabilities are evaluated, including asset management, safety monitoring and emergency response capabilities.

Indicators include safety testing, safety operation and maintenance, safety disposal, safety inspection, and safety audit.

(3) Security effect indicators. Mainly the actual operation effect of the smart city security operation system.

Evaluation is conducted, which includes indicators such as security vulnerabilities, security incidents, and attack and defense confrontation.

(4) Intelligent operation indicators. Mainly the degree of integration of smart city AI + security operation system

Evaluation will be conducted, including the security level of the AI capability platform, AI compliance, actual effects, etc.

6. AI empowers smart city security solutions

6.1. AI+Smart City Cybersecurity

Smart city AI+ network security is to protect and enhance the security of the network layer through artificial intelligence technology It covers technologies such as next-generation firewall, full-flow threat detection, intelligent routing and load balancing, and Unlike traditional network security, AI+ intervention can solve problems that are difficult to solve in traditional security fields, such as encryption and tunneling. Dao technology provides a solution.

6.1.1. Next-generation firewall

Smart cities will widely deploy firewalls to deal with increasingly complex and diverse network threats. The next-generation firewall with AI technology support not only has the packet filtering function of traditional firewalls, but also Through artificial intelligence and machine learning, it is possible to analyze network behavior, identify abnormal patterns, and combine Threat intelligence feeds provide real-time threat feedback and response. These features enable it to provide Provides more comprehensive and intelligent network security protection than traditional firewalls, effectively coping with modern network environments multiple threats.

6.1.2. Full-flow threat detection

In smart cities, encrypted traffic accounts for more than 70% of network traffic. Traditional threat detection methods AI full-flow monitoring uses the best supervised machine learning algorithms and network protocol restoration technology to technology, training classification tags and establishing an incremental learning database, building an automated threat detection system, and implementing The system uses a supervised machine learning algorithm to perform sample This collection, processing, model training, and verification use leaf-wise splitting strategy for classification to improve detection The accuracy is improved and the model training time is shortened. At the same time, the mutually exclusive feature bundling algorithm is used to improve the detection efficiency. Reduce memory consumption and effectively detect advanced threats such as malicious encrypted traffic, DGA domains, and covert tunnels.

Combine threat intelligence and detection models to comprehensively identify threat behaviors in network traffic.

Provide support for smart city security operations.

6.1.3. Intelligent routing and load balancing

Intelligent routing and load balancing in smart cities are the key to ensuring efficient, secure,

AI is an important technology for reliable operation. It can optimize network routing and dynamically adjust path options to avoid congestion.

and potential security threats. For example, security policy integration, adaptive routing, machine learning real-time optimization routing

AI-based load balancing prediction analysis to ensure efficient use of network resources and prevent single points of failure

With the support of AI, these technologies can more efficiently manage the network infrastructure in smart cities.

Ensure reliability and performance of all types of urban applications.

6.1.4. Security Intelligent Question and Answer

The existing security intelligent question-and-answer application is developed based on the knowledge graph combined with the long short-term memory network model.

The answering mode is fixed, the overall ability is limited, and the security knowledge question-answering based on AI and security big model is

The technology introduces secure corpus, which not only takes advantage of the human-like ability of large-model NLP natural language, but also

The professionalism of the model’s answers is ensured. In addition to real-time response scenarios, the model also

In smart cities, good

Interactivity can provide support for various security tasks and improve the usability of various systems.

6.1.5. Threat Intelligence Analysis

Threat intelligence analysis introduces artificial intelligence and machine learning technologies to timely discover, analyze and respond to potential threats.

It can provide a higher level of security for smart cities by preventing cyber attacks and security threats.

Threat intelligence analysis technology avoids the merging and deduplication based on preset data labels, and combines manual operation

The traditional way of the camp solves the problems of large amount of intelligence data, strong heterogeneity and poor timeliness, and

In smart cities, system operation logs, user behavior logs, traffic data and external public intelligence can be collected.

Data, hacker forums, and other data sources are automatically labeled, and machine learning and deep learning are used to

It uses intelligence aggregation and classification technologies to conduct security risk assessment,

Provide capability support for early warning and security incident response.

6.1.6. Automated penetration testing

Penetration testing can effectively assess the security status of systems in smart cities and propose reasonable improvements

The traditional penetration test work is highly dependent on the human judgment and operation of security personnel.

There is room for improvement in both the efficiency and the cost. Automated penetration testing based on AI big models combines models and tools

The big model has the input and output of penetration testing and the feedback capability of process reasoning, and cooperates with various automatic

Automated scripting and intelligent process applications allow systems in smart cities to automate penetration testing.

Assist penetration test participants, provide efficient workflow and generate reliable reports.

6.2. AI+Smart City Application Security

Smart city AI+ application security is a new defense technology that acts on the application layer.

In these applications, artificial intelligence technology is used to ensure that applications can run safely, prevent data leakage,

Defend against cyberattacks and protect the privacy of citizens and the integrity of city infrastructure.

Control, mimic honeypot, and content detection are new technologies based on AI.

6.2.1. Risk Control

The risk control system uses AI technology to build a dynamic authentication mechanism for business access and access behavior analysis.

Analysis, the pursuit of actively building an ever-changing access environment, so as to achieve automated attacks and unknown

Risk defense, effectively protecting against machine behaviors such as crawlers and AI script attacks. Dynamic intelligent identity authentication

By generating intelligent dynamic tokens and dynamic fingerprints, it continuously tracks and analyzes access sources and behaviors to confirm access

Intelligent behavior analysis uses multi-dimensional feature collection and in-depth

Learn, train and optimize models to identify threat behaviors. This defense approach also makes up for

The defects of traditional WAF greatly reduce the possibility of successful attacks. At the same time, the risk control system can

Combined with traditional detection rules, it provides comprehensive dynamic and static defense for smart city Web applications.

6.2.2. Mimic Honeypot

Mimic honeypots lure attackers into the network by simulating real targets, such as operating systems and network devices.

AI honeypots can better analyze attack behaviors.

The AI-driven analytics platform uses machine learning technology to automatically

Identify attack methods and features, thereby establishing a deception system to effectively discover and study attack behaviors.

The technology enables the rapid connection and deployment of honeypots, and the establishment of deception grids and the placement of baits in business environments.

The honeypot system can intelligently establish a deception defense system and simulate different environments to trigger attacks through attack method analysis.

Once an attacker breaks in, the system will issue a high-precision alarm and

AI technology is used to clean and denoise data, clarify attack intent, and conduct associated warnings and defense strategies.

At the same time, AI technology can realize event-driven automatic response and automatically create simulation deployment.

6.2.3. Content security governance

A method of preventing unauthorized, malicious or inappropriate information from being transmitted by analyzing and filtering application data content.

Information spreads through the network. AI can identify risky content features more quickly by learning malicious content.

By building a content risk control system, we can give full play to AI's capabilities in semantic understanding, image recognition, audio recognition, etc.

The advantages of AI big models are used to pre-train text, audio and video data review capabilities.

The improvement of strength.

6.2.4. Supply Chain Security Intelligent Analysis

The software supply chain has become the focus of the industry, covering the entire process of software development from source code review to delivery and deployment.

The entire deployment process ensures that the software is safe and reliable at every step.

In the smart city, supply chain information maintenance, open source software management, closed-loop software materials

can be processed by large models, and third-party open source libraries and binary packages can be used without decryption.

In addition, the use of big security models to build software can be done from development to operation.

The entire process of executing security testing tasks includes strategy generation, task execution, result collection and report summary.

In terms of conclusion and other aspects, one-stop, intelligent smart city software supply chain security risk management can be achieved.

6.2.5.API Security Intelligent Monitoring

The API of business systems has become a key target of hacker attacks. Existing API security solutions can only

Insufficient management capabilities often lead to data leakage, malicious attacks, unauthorized access, business operation failures, etc.

Through AI and big model technology, API assets in smart cities can be protected.

Automatic discovery and identification technology, establishment of API asset management library, monitoring and analysis of API abnormal operations, identification

Many technologies will be used in smart cities to prevent hidden threats (illegal use of tokens, data leakage and loss, etc.)

It plays a key role in the functional interface, combines with AI big models to provide risk management suggestions, and enhances the API in use.

Security prevention capabilities during the use and content provision process.

6.2.6. Malicious code detection

Through static feature analysis (such as checking the structure of binary code) and dynamic behavior monitoring (such as

Such as tracking the sequence of operations at runtime), AI technology has learned and mastered the behavior of past malware.

This allows the AI to identify new, previously unseen malicious files, especially

It is very effective in identifying zero-day attacks. In addition, with continuous training, AI can gradually improve

The accuracy of identifying malicious files can reduce the possibility of false positives and false negatives, and can also quickly adapt to malicious

New variants of malicious software.

6.2.7. User and entity behavior analysis

Smart cities provide a wide range of services and have many users. Relying on the reasoning ability of large models, innovative user access

Abnormal behavior analysis model. Automatically complete the analysis of network logs, user behavior logs, and database logs.

Output highly accurate abnormal behavior warning information. Build an abnormal behavior detection model through deep learning algorithms.

It can predict and identify unknown abnormal behaviors, and assist security experts in carrying out data security operation services.

6.3. AI+Smart City Data Security

Based on AI+ data security capabilities, it is responsible for protecting various types of urban data in smart cities, including AI data

Security covers multiple dimensions, from data storage to model training and deployment, to real-time monitoring and response.

Should. Use data control platform, AI data watermark and other technologies to ensure the security and reliability of privacy.

6.3.1.AI Data Watermarking

It is a technology used to protect data ownership and ensure data integrity.

By inserting imperceptible but verifiable identifiers, data watermarking technology helps monitor data usage and prevent

Prevent unauthorized copying and tampering, and protect the copyright and security of data. It is worth mentioning that with the production

With the development of AI, data generated by AI also needs to be identified and distinguished through watermark technology to prevent data from being leaked.

data confusion and attack.

6.3.2. AI data classification and grading

Enhance the relevant laws and regulations, industry standards, corporate norms and business knowledge based on important data identification

The general large language model has the ability to analyze and reason in specific vertical fields, combining structured, semi-structured, and non-structured

Convert structured data sources into natural language to build intelligent identification of important and core data

And classification and grading labeling technology, so as to improve the automation, accuracy and efficiency of data recognition.

6.3.3.AI Data Security Audit

Using artificial intelligence technology to combine supervised learning and unsupervised learning, we can establish a flexible

Intelligent auditing technology can learn from the user's historical behavior data to form a user's specific

In the subsequent audit process, the trained behavior

It is a baseline model that can monitor and detect in real time whether the user's current behavior deviates from the normal behavior range.

In this way, abnormal business operation behaviors can be discovered, and the relevant data of audited historical alarm events can be learned.

Through learning and training, we build intelligent audit models to automatically identify whether the alarm data contains security incidents.

documents and automatically audit them.

6.3.4.AI Data Security Compliance Tools

AI-based data security inspection toolbox to achieve intelligent and automated assessment of data security

Use NLP to build an intelligent evaluation matrix model to automatically analyze user needs and generate target

The evaluation matrix can be used to interact with users in natural language and understand their needs and problems.

Automatically create and execute packages based on the company's existing assessment information, business processes, security measures and other related information

An evaluation matrix containing inspection points, inspection questions, and materials required for inspection.

Evidence is deeply analyzed, matching assessment and inspection information is automatically

Evaluate the analysis results and determine the degree of satisfaction of the evidence and inspection information.

6.3.5.AI Data Desensitization

AI technology can realize dynamic information desensitization in the process of using data, which means

The displayed data format is automatically adjusted according to the user's identity, access environment and usage scenario.

If the user is an internal auditor, more details may be required; for external cooperation

partners, we will display a highly de-sensitized view of the data in order to maximize the

Minimize the risk of data leakage.

6.4. AI + Smart City Public Security

The close integration of smart city construction and AI has brought revolutionary changes to the field of public safety.

Through artificial intelligence technology, real-time monitoring and analysis of the urban environment can be achieved, and pollution sources can be identified.

Accurately locate and issue early warning information in a timely manner, providing a scientific basis for environmental protection decision-making.

Governance and environmental optimization have improved the quality of the urban environment. In terms of public safety, AI technology applications

In terms of face recognition, video surveillance, event warning, etc., it helps cities achieve precise safety control.

Through functions such as face recognition and license plate recognition, it assists in public safety management and public security maintenance, and realizes

Intelligent identification and real-time warning of emergency events improve the efficiency of emergency rescue work and effectively ensure public safety

Complete.

6.4.1. Social Governance Security Plan

Smart city social governance security solutions can use video surveillance and intelligent video analysis technology to monitor

Control key areas, prevent criminal activities, and maximize public safety.

Intelligent traffic management with AI algorithms not only optimizes traffic flow and reduces accidents, but also responds quickly to emergencies.

In addition, the smart patrol robot provides 24-hour service, combined with artificial intelligence face recognition and other technologies.

technology to ensure community safety and improve public security. It cannot be ignored that predictive policing uses data analysis to

The comprehensive application of these technologies enables cities to

Security management is smarter and more efficient.

6.4.2. Disaster monitoring and early warning plan

Disaster monitoring and early warning are the key to smart city disaster prevention and mitigation, including artificial intelligence technology

Disaster warning systems, emergency command and dispatch, and intelligent firefighting systems.

Detection equipment and flood warning systems, combined with large model algorithms, disaster warning systems can predict natural disasters in advance

to minimize the damage caused by disasters.

The emergency command platform can coordinate the resources of various departments to achieve rapid response and processing.

The system monitors fire hazards through devices such as smoke and temperature sensors, and can promptly notify relevant departments to

The three together constitute a comprehensive disaster monitoring and early warning program, aiming to prevent natural and man-made disasters.

Disasters and protect people's lives and property.

6.4.3. Public health and safety program

The public health safety program mainly includes three aspects: disease monitoring and early warning, food safety management and environmental monitoring.

By using big data and AI technology, the disease surveillance and early warning system can monitor and predict diseases in real time.

To ensure food safety,

Traceability systems and sensors closely monitor food production and supply chains. Environmental monitoring systems can

Regularly monitor environmental indicators such as air quality, water quality, and noise, and handle any abnormalities promptly to prevent environmental

This comprehensive public health and safety program can effectively prevent

Control all kinds of health risks and ensure the safety of public life.

6.4.4. Safety production management plan

Safety production involves industrial Internet security and operating environment monitoring, which requires the construction of artificial intelligence-based

Industrial control systems for monitoring and management. Operating environment monitoring is achieved by using sensors and AI computing platforms.

Real-time monitoring of toxic and harmful gases and noise levels in the workplace to ensure the health and safety of workers.

7. China Mobile Smart City Artificial Intelligence Security Reference Case

7.1. Case Study on Risk Prevention **of AI** in Smart Cities

7.1.1. Security protection of Heilongjiang Provincial Ocean Computing Government Affairs Big Model

(1) Background and needs

Since the launch of the Heilongjiang Province Digital Government Project in 2022, it has adhered to data-driven and innovative

Leading and continuously accelerating the application of artificial intelligence. With its rich experience in building digital government,

China Mobile uses a general large model as the basis and integrates data from the government sector for fine-tuning.

The government affairs constraint model was introduced to restrict the output, and finally successfully created a

The industry big model - Jiutian Haisuan government affairs big model.

The Haisuan government affairs big model deeply integrates "government affairs policy-government affairs-government affairs data storage" into the model.

By issuing natural language instructions to the large model,

Access deep databases, connect multiple sources, complex and heterogeneous data tables, and quickly obtain intuitive data

Analysis results. Currently, we have trained more than 10 types of data, 400 billion private domain data, and 100,000 precision data.

In terms of security, the Haisuan Government Affairs Big Model uses the government affairs expertise in the information field to

The model is enhanced by course learning and aligned generalization, while coordinating private domain data as the final result.

Through the dispatching capability of the government information field, scattered related data are gathered to focus on user consultation.

To solve all problems in the field; to expand the boundaries of government services and realize proactive services;

Without leaving the venue, you can achieve reliable responses to government affairs issues and ensure that government services are safe and controllable.

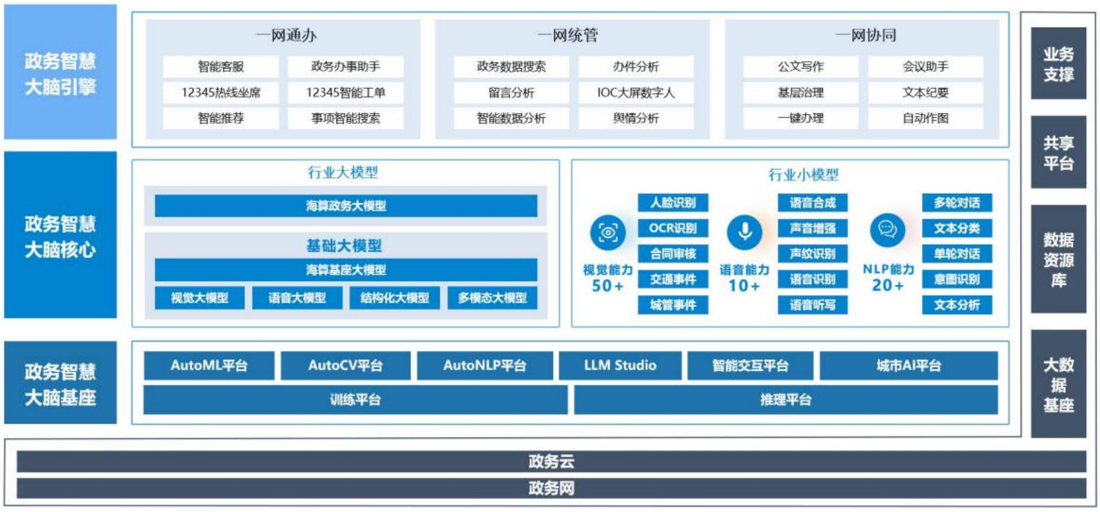


Figure 5: Construction plan of Haisuan government affairs big model

(2) Construction plan

Based on the principles of demand boundaries, business boundaries, and security boundaries, we will conduct four-dimensional analysis on cloud, network, data, and usage.

Conduct detailed research and confirmation to ensure the implementation of 26 security capabilities during the construction process.

Modules such as display, capability output, business support and data aggregation. Supporting construction of external docking platform, security

Full operation management system, safe operation service system, safe operation physical environment, etc. Joint protection

For the provincial government big model platform, through cloud, network, data, and application security operations

The project implemented comprehensive and strict network and data security protection measures, combined with effective cooperation

The project team has set up a

The platform's application monitoring system tracks and analyzes application behavior and related

Data flows, and possible problems are discovered and solved immediately.

Regarding training data management, this project defines a strict process.

Training with properly cleaned and anonymized data ensures data security and privacy from the source

At the same time, we classify and label the refined training data to ensure that we can

Clearly know which data is used for what purpose. The project team regularly conducts data quality checks and updates.

New to ensure that large models are trained based on safe and accurate data.

Through such a series of data and network security protection measures, including but not limited to access control,

Network encryption, data monitoring, vulnerability scanning, compliance management, application monitoring, and training data management,

We have built a secure, stable and reliable Haisuan government affairs big model platform to provide users with

A high-quality usage environment.

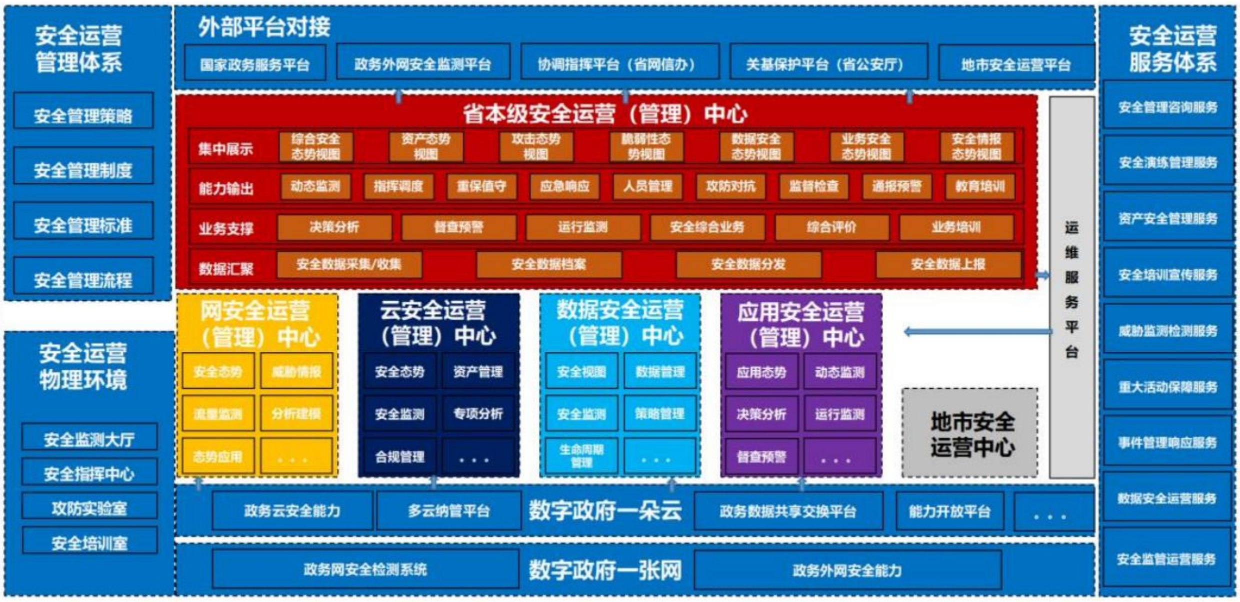


Figure 6 Heilongjiang Provincial Smart City Government Affairs Security Construction Plan

(3) Construction achievements

The Haisuan government affairs big model has built 12345 intelligent hotline, government affairs intelligent search, government affairs intelligent assistance

It has four application scenarios, including the digital government project of Heilongjiang Province, the

Government affairs model project, Guangdong Province Joint Laboratory case, Shenzhen People's Livelihood Appeal (12345 hotline)

Projects, etc.

The security capability construction of the Haisuan government affairs big model continues to advance, from cloud, network, data, and application.

In addition, the company also protects against technical risks of large models through compliance management, training data management, content management,

The compliance requirements of large models have been met. So far, no security-related incidents have occurred and the system is stable.

Stable operations provide customers with quality services.

7.1.2. China Mobile AI Model Vulnerability Assessment Platform

(1) Background and needs

The application of large artificial intelligence models is becoming more and more widespread, facing the threat of attackers using jailbreak attacks and target hijacking.

By means of holding and prompting leakage, the defense strategy of the artificial intelligence big model is bypassed to illegally obtain big data.

The sensitive information of the model poses a huge security risk of industry knowledge being stolen.

They also affect the accuracy of the output of the large model by poisoning the fine-tuning data, or by

By inputting adversarial samples into the large model, the large model is induced to make incorrect inferences.

Various components and middleware used to build large models may have software vulnerabilities, which may cause large model parameters to

Data or large model applications are stolen or illegally controlled, causing abnormalities in smart city smart applications.

Based on its own experience, China Mobile launched an AI model evaluation platform to solve the problem of artificial intelligence

Model vulnerability assessment.

(2) Construction plan

In response to the above requirements, the construction plan is shown in the figure.

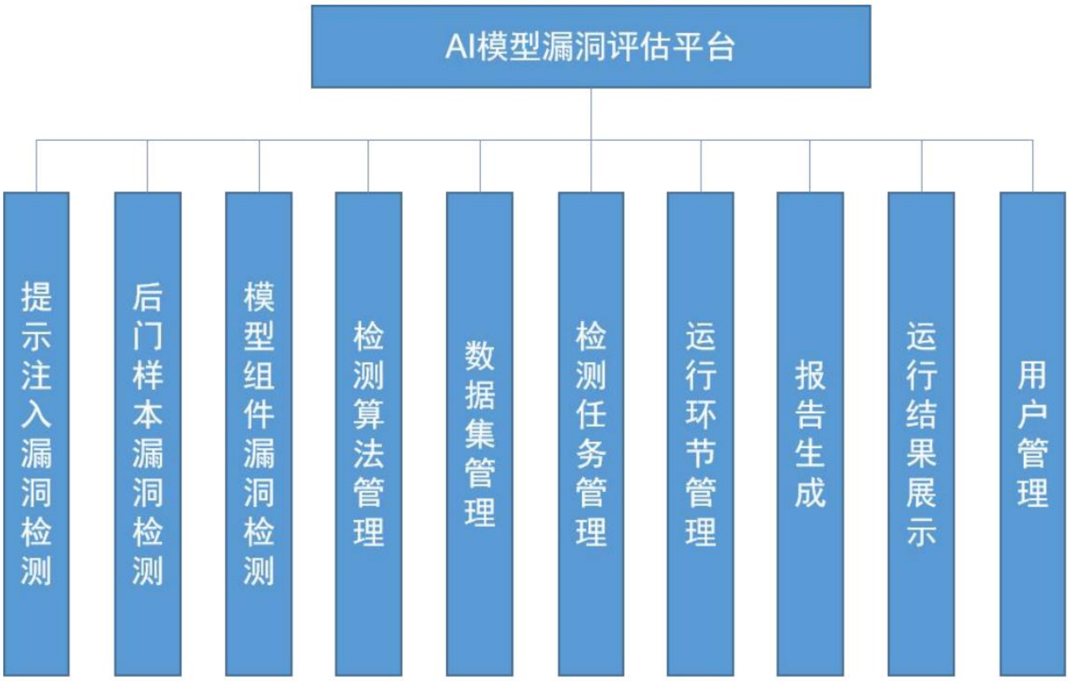


Figure 7 AI model vulnerability assessment platform architecture diagram

The Big Model Security Assessment Tool is a comprehensive and efficient vulnerability assessment tool designed to ensure that

The system uses reinforcement learning and deep learning to ensure the safety of the model in various application scenarios of smart cities.

Through reverse deduction and interactive verification of large models, possible injection indicators can be detected.

and backdoor instructions, and conduct security analysis on the large model framework and related components, thus finding that the large model

Security vulnerabilities and risks that may exist during component calling, model loading, algorithm running, etc.

The system also provides a wealth of security reports and visualization tools to help users gain a deeper understanding of the

The safety status of the model provides strong support for the optimization and reinforcement of large models.

The tool evaluation dimensions include model poisoning, prompt injection, and sample adversarial aspects, and generates large models

Safety assessment report.

Tip injection vulnerability detection

Model prompt injection vulnerabilities can cause attackers to use the vulnerabilities to construct malicious prompts to interfere with the model

The prediction results of the model cause the model to output wrong or unpredictable information. Prompt injection vulnerability detection module

By analyzing the model and its input, we can compare the model's output under normal input and malicious prompts.

Differences can be used to identify possible hint injection vulnerabilities. When the system detects a vulnerability in the model,

Propose corresponding evaluation results and repair suggestions to improve its defense capabilities against malicious prompts.

Backdoor sample vulnerability detection

When the model backdoor sample vulnerability is exploited by an attacker, it may cause the model to

The expected wrong output is generated when the input is input. The existence of this vulnerability makes the model more vulnerable to attackers.

The model backdoor sample vulnerability detection module uses the characteristics of the backdoor sample to detect

Perturb the model input and observe whether the output changes to determine whether the model has a backdoor pattern.

When the system detects a vulnerability in the model, it will present the corresponding evaluation results and repair suggestions.

to exclude possible backdoor samples, so as to prevent these maliciously injected data from affecting the model Accuracy and credibility.

Model component vulnerability detection

Model component vulnerabilities are the vulnerabilities that may exist in the components used in the model training and application process. Security flaws or misconfigurations may lead to attacks on models, data leakage, performance degradation, and other issues. Risks such as decline. Model component vulnerability detection module source code analysis technology and vulnerability scanning technology Automatically detect model components to find possible vulnerabilities and security issues, and then analyze The severity and potential impact of detected vulnerabilities, assessing the security and reliability of model components, Generate corresponding repair measures and suggestions based on the evaluation results to improve the safety and stability of model components Qualitative.

Detection algorithm management

The detection algorithm management module is responsible for managing and maintaining the prompt injection vulnerability detection algorithm and backdoor samples Vulnerability detection algorithm and model component vulnerability detection algorithm. Provide algorithm storage, version management, parameters Preset and interface configuration functions to ensure the availability and consistency of the detection algorithm. Internal algorithm interface management makes it convenient for users to add new algorithms and configure new detection strategies.

Dataset Management

The data set management module is responsible for collecting, organizing, storing and managing typical data sets used for testing. Provides rich and reliable data basis for the detection module. In addition, the data set management is also responsible for external data Access and save to facilitate the detection module to detect external data.

Inspection task management

The detection task management module is responsible for creating and executing detection tasks, The detection task management module detects vulnerabilities, backdoor sample vulnerabilities, and model component vulnerabilities.

By automatically scheduling vulnerability detection algorithms, the system can efficiently and accurately detect whether the model exists

When a vulnerability is found, the system will alert the user to the vulnerability event, and then the module root

According to the security strategy formulated by the user, the vulnerability is marked and improvement suggestions and solutions are generated, so as to effectively

Prevent vulnerabilities from affecting model performance.

Operating environment management

The operating environment management module is the basic support of the system, which is responsible for managing and maintaining the operation of the algorithm.

environment to ensure the availability of the algorithm. At the same time, the running environment management module also provides

The algorithm provides the basic environment and adaptive environment deployment capabilities to ensure the normal operation of the algorithm.

Report Generation

The report generation module can automatically generate detailed test reports based on the test results, providing users with

Provide comprehensive analysis of test results.

Operation results display

The operation result display module is responsible for displaying the test results to the user in an intuitive way, which is convenient for

User analysis and understanding.

User Management

The user management module is responsible for user information registration, login, authority management and other operations to ensure the system

System security and user rights.

(3) Construction achievements

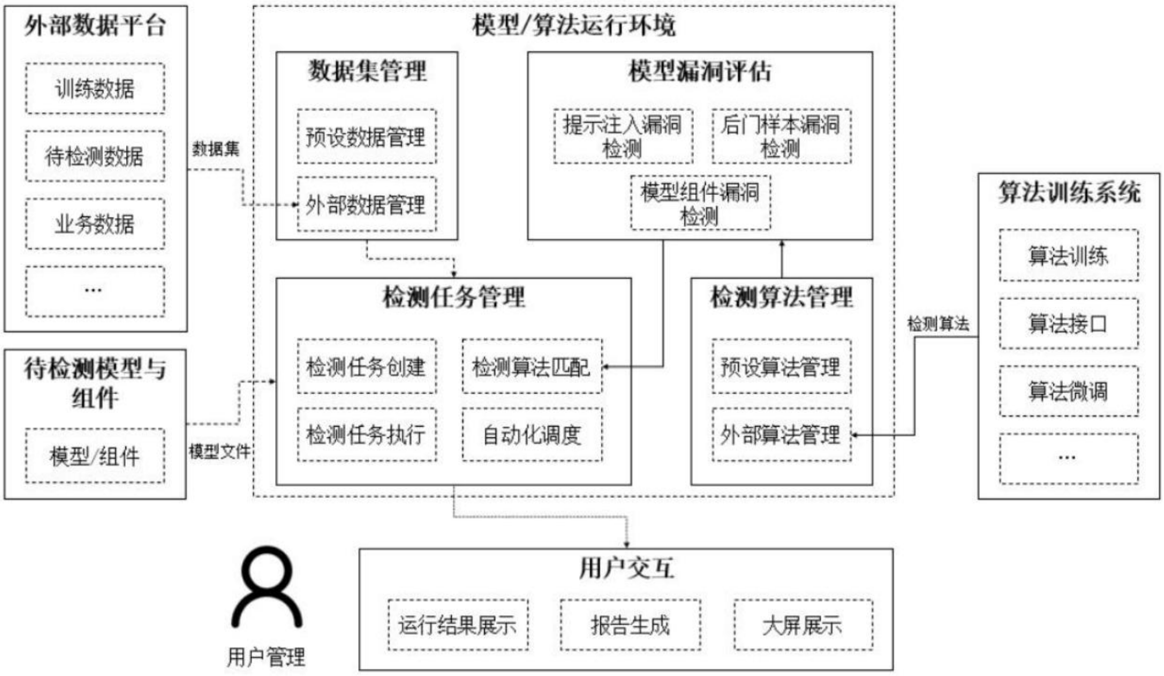


Figure 8 Smart City Big Model Security and Compliance Solution

The large model security assessment tool function has been applied in the smart city supporting industrial large

The initial application of the model and financial big model has been implemented in the online evaluation. So far, the industrial big model has not been

Regulators report and detect compliance, network and data security risks, and protect user privacy and

Data security; in the future, a series of supporting capabilities such as API security monitoring will be added in sequence.

Many rural commercial banks and property insurance companies in Panjin, Liaoning, Hefei, Anhui, Wenshan, Yunnan, etc. provided security

Stable risk management assistance service. Currently, the large model security capability enables more than 20 applications.

This tool improves the integrated security protection system for industrial and financial large models, ensuring that

The large-scale industry and financial underwriting model was launched safely.

7.2. **AI-** enabled smart city security cases

7.2.1. Case study of the construction and operation of Zhongshan Municipal Government Information Security System

(1) Background and needs

In June 2020, the Zhongshan Municipal Government Services and Data Management Bureau built a

However, with the rapid expansion of government service scenarios, the security operation and maintenance management center for the flow of police and work orders has emerged.

The current situation of alarm fatigue, expert shortage, and efficiency bottleneck has been solved.

With the rapid development of deep learning and machine learning technologies, AI big models are playing an increasingly important role in image recognition,

These technological advances have greatly promoted social productivity.

The improvement of the environment and the convenience of life are accompanied by a series of safety risks and challenges.

The cost of hacker attacks and learning is getting lower and lower, which leads to greater risks to government network security.

The increasing number of automated attacks has led to an increase in the number of alerts that the government needs to deal with, so security experts are needed.

The model is introduced to solve the security issues brought about by new technologies.

By supplementing the network security big model capabilities, asset vulnerability management, threat detection and discovery, and alarm

The capabilities of event closure, investigation, summary and reporting have been improved, and artificial intelligence technology continues to empower threats

Detection and situational awareness, terminal, border, cloud security and other security products constitute the foundation of efficient and intelligent operations.

Security operation and maintenance center.

(2) Construction plan

By deploying a large security model, the traffic, logs, and codes of the government extranet security operation and maintenance center can be connected.

Data pre-training, traffic understanding ability, code understanding ability, attack and defense understanding ability and security common sense

Identification and understanding capabilities to achieve asset vulnerability management, threat detection and discovery, alarm event closure, disposal investigation,

Improve the ability to summarize and report.

1. Security data analysis and security alarm assessment: Supports multi-source and multi-type threat data aggregation and noise reduction.

The warning features after noise reduction are correlated to find threat events, and the scope and formation of threat events are analyzed. Cause, attacker identity and risk assessment, etc., support natural language-based interpretation of attack traffic, Analyze attackers' methods and intentions, and discover the capabilities of comprehensive security situation analysis and security big models. Unknown threat.

2. Security incident response and security knowledge Q&A: Support automated clue evidence collection and analysis of alarm events, Statistical analysis, traceability analysis, assistance or active blocking, and control of security equipment. The ability to deeply understand and process text data in the security field requires accurate parsing and analysis of security-related Text information and respond to user needs based on contextual statements.

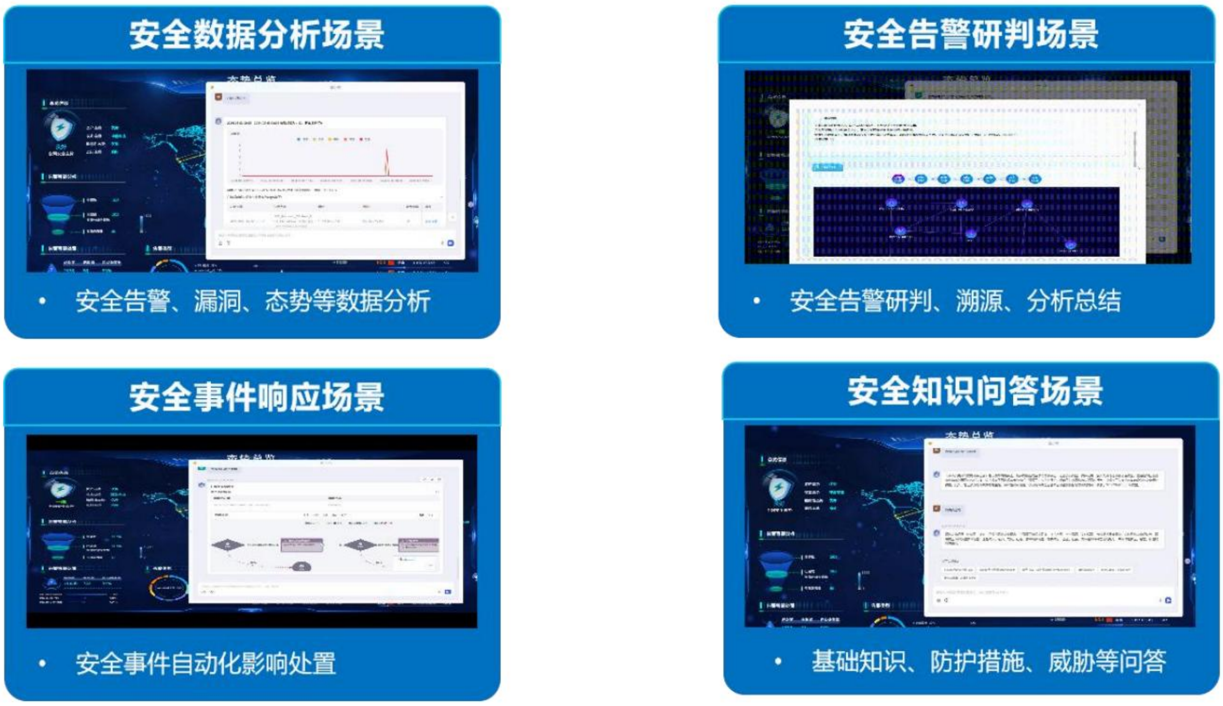

安全数据分析场景
• 安全告警、漏洞、态势等数据分析
安全告警研判场景
• 安全告警研判、溯源、分析总结
安全事件响应场景
• 安全事件自动化影响处置
安全知识问答场景
• 基础知识、防护措施、威胁等问答

Figure 9: Case study of government information security construction in Zhongshan City

(3) Construction achievements

This project is based on the security operation and maintenance of the Zhongshan Municipal Government Services and Data Administration in June 2020. In the management center, the network security big model is combined with security operation and maintenance for the first time, covering "cloud, network, data, The integrated security system architecture of in-depth defense at the terminal level has comprehensive risk perception, real-time monitoring and early warning, The ability to conduct accurate analysis and rapid response to security operations ensures the safety of the city's government information system.

The project covers the Political and Legal Committee of Zhongshan Municipal Committee, the Justice Bureau, the Transportation Bureau,

26 units including the Transportation Bureau, Agriculture Bureau, and Market Supervision Bureau have occupied a strong position for subsequent business expansion.

form.

7.2.2. Operational Case of Venusstar Anxing Intelligent Security Assistant

(1) Background and needs

Compared with traditional network security methods, artificial intelligence has unique advantages in network security.

Artificial intelligence can learn and identify patterns from large amounts of data through machine learning and deep learning techniques.

Uncover hidden correlations and anomalies, and AI can quickly adapt to new threats and provide real-time

Therefore, it is necessary to continue to research and develop artificial intelligence technology to strengthen its role in network

Applications and innovations in security to ensure the security and trustworthiness of cyberspace.

Anxing Intelligent Security Assistant is a full-scenario intelligent security operation assistant based on the security big model.

Natural language command understanding, security data analysis and interpretation, intelligent decision-making and automatic security operation tasks

It is the decision-making and control core of the new generation of intelligent security operation platform.

Based on the characteristics of the industry, Anxing Intelligent Security Assistant has developed a unique security vertical field based on AI security intelligence.

The "large and small models autonomous collaboration" technology system combines Venustech's deep security product capabilities and AI expertise

The accumulation of small security models and the integration of the powerful intention understanding and reasoning capabilities of the large model have built a

A full-scenario security intelligent automation operation center driven by a large security model.

Collaboration" technology system, Anxing Intelligent Security Assistant can serve as the entrance to security operations, which can greatly reduce

The technical threshold of security operation is improved, the execution efficiency of security operation is improved, and accurate

The next generation of intelligent security operation system that drives security products and special security models to perform security operation tasks

Test.

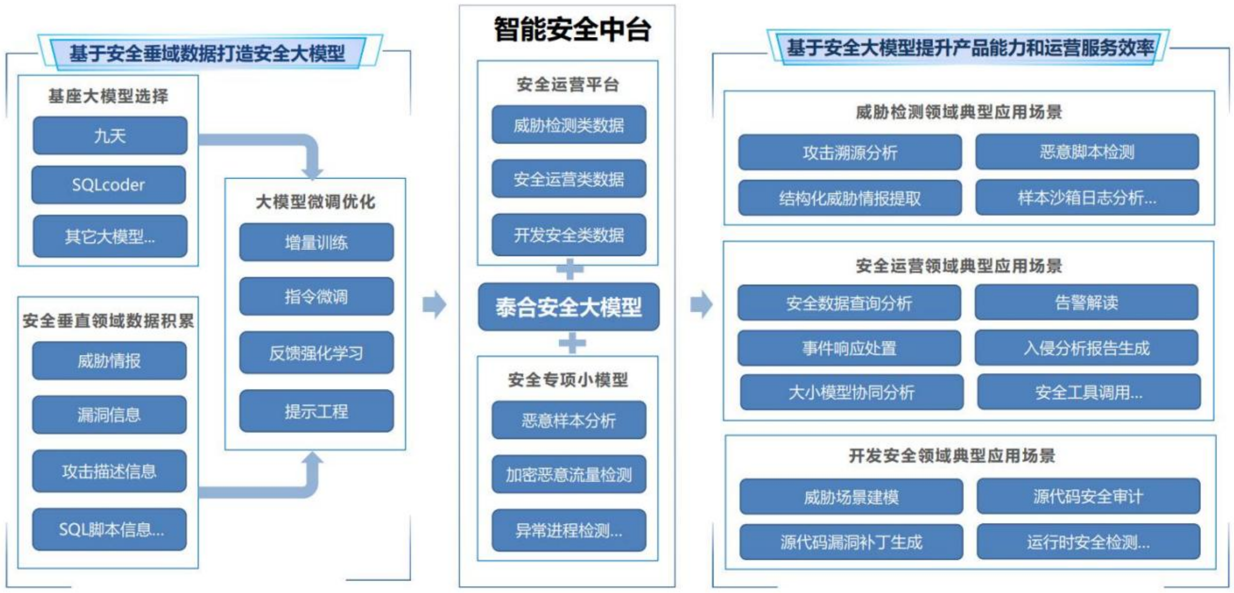


Figure 10 Anxing Intelligent Security Assistant Technology Application Route

(2) Construction plan

Smart Center: The smart assistant uses natural language interaction as the core to provide multi-person collaborative processing of security incidents.

The interactive platform for security configuration, security command issuance, and script execution enables

In the interactive interface, complete the interaction with security personnel, security personnel issue security instructions, and call security scripts

The collaborative robot can respond 24/7 and can perform intelligent image recognition and support image text

Extract word content, assist in issuing or parsing security instructions, and intelligently recommend relevant knowledge bases and handling actions.

Improve incident response efficiency.

Cybersecurity domain knowledge services: Use large model full parameter pre-training and supervised fine-tuning to

The model injects security vertical domain knowledge and uses the retrieval-augmented generation (RAG) technology to flexibly inject security vertical domain knowledge into the large model.

Type injection includes the latest threat intelligence, network knowledge and product manuals, etc.

Data analysis and presentation: Rapidly display data through natural language interaction to help users quickly understand the security

Comprehensive situation of all alerts, risky assets, risky users, etc.

Connect with security products and capabilities: More than 100 products and capabilities have been connected to support safe operation.

System scheduling of multi-brand and multi-type security products commonly seen in the camp to realize the arrangement of security operation tasks and scheduling execution.

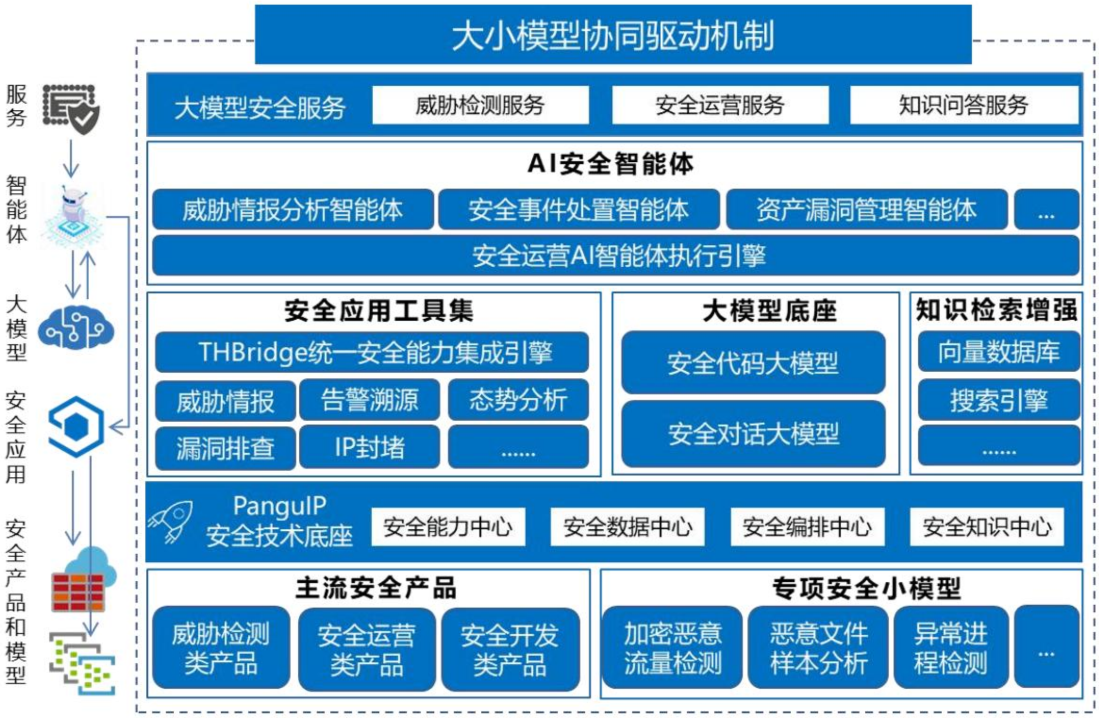


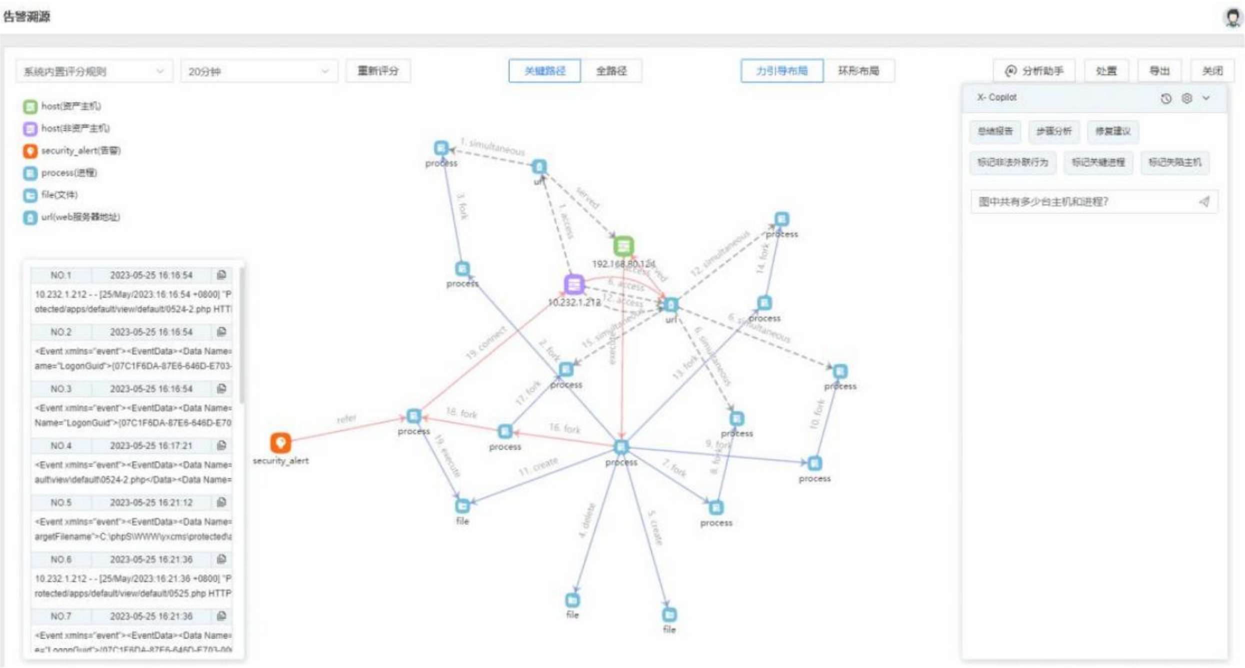
Figure 11 Anxing Intelligent Security Assistant Technical Architecture

Automated analysis, focusing on key alarms: Organize two types of traceability diagrams, one is alarms without important clues

The traceability graph of indicates that the attack was unsuccessful or a false positive. The attacker did not perform any operation on the victim host.

There are no further clues; second, the traceability diagram of the alarm with a clue evidence chain indicates that the attack was successful.

The victim's host generates file and process related operations, leaving clues.



```
NO.1    2023-05-25 16:16:54
10.232.1.212 - - [25/May/2023:16:16:54 +0800] "P
otected/apps/default/view/default/0524-2.php HTT
NO.2    2023-05-25 16:16:54
<Event xmlns="event"><EventData><Data Name=
ame="LogonGuid">{07C1F6DA-87E6-646D-E703-
NO.3    2023-05-25 16:16:54
<Event xmlns="event"><EventData><Data Name=
Name="LogonGuid">{07C1F6DA-87E6-646D-E70
NO.4    2023-05-25 16:17:21
<Event xmlns="event"><EventData><Data Name=
ault\view\default\0524-2.php</Data><Data Name=
NO.5    2023-05-25 16:21:12
<Event xmlns="event"><EventData><Data Name=
argetFilename">C:\phpS\WWW\yxcms\protected\a
NO.6    2023-05-25 16:21:36
10.232.1.212 - - [25/May/2023:16:21:36 +0800] "P
rotected/apps/default/view/default/0525.php HTTP
NO.7    2023-05-25 16:21:36
<Event xmlns="event"><EventData><Data Name=
e="LogonGuid">{07C1F6DA-87E6-646D-E703-00
```

Figure 12 Anxing Intelligent Security Assistant Alarm Tracing Diagram

Closed-loop management of vulnerability issues: calling the engine to scan target assets and obtain vulnerabilities, configurations, and weak points

According to the latest vulnerability intelligence, match the vulnerabilities in the asset library

The asset system type, version, port and other information can be used to identify assets that may have vulnerabilities.

Compare historical scan results with disposal status, track vulnerability disposal status in real time and generate vulnerability summary

Report, dispatch disposal work orders to asset responsible persons

Intelligent scheduling platform, security products and tools to respond: with mining virus processing, ransomware

Virus processing, host outbound attack event processing, (compromised host), phishing emails, honeypot threat hunting

Attack protection, one-click search for people and assets, ARP hijacking disposal, DLP linkage database desensitization disposal solution,

Bastion machine bypass disposal, DDOS attack disposal, user operation and maintenance authority management automation, abnormal DNS

Seeking related capabilities such as automated analysis and one-click blocking.

(3) Construction achievements

In a provincial smart city project, security operations were carried out for more than 30,000 IT assets.

Anxing Intelligent Security Assistant can help customers save 375 man-days of manpower per month and 4,500 man-days per year.

The itinerary features the following achievements:

Centralized dispatch: Security platforms and equipment are centrally dispatched to form a united front to assist in security operations.

Rapid response: Rapidly respond to massive security operations in parallel through intelligent security operations systems.

All-weather duty: 7×24 hours×365 days, uninterrupted safety operation duty;

Labor, reverse the unfavorable situation of offense and defense.

Intelligent operation: The daily work of security operation is handled through intelligent operation tasks, and the work progress is transparent and

See, assisting users to track the closed loop.

Security experts: Provide common security analysis and security expertise Q&A capabilities to reduce security operations

The experience gap between personnel can be reduced and the overall operational strength of the security operation team can be improved.

8. Outlook on China Mobile's AI+ Smart City Security

8.1. **AI** makes smart cities safer

8.1.1. Improve the legal system and safety standards

At present, my country's existing laws and regulations are being gradually improved.

The promulgation and implementation of the "Regulations on the Administration of Generative Artificial Intelligence" and the "Regulations on the Administration of Generative Artificial Intelligence" will ensure the safety,

Transparency and ethics, while also promoting technological innovation and protecting personal privacy.

Existing laws and regulations provide important guidance for city managers in the general direction of the safe development of artificial intelligence.

Guidance has been provided to ensure that smart city AI has laws to follow and that work in related fields can be better organized.

Smart city managers should strictly abide by national and local laws and regulations and implement smart city artificial intelligence

Management related work.

The national standard system for "Smart City Artificial Intelligence Security" still needs to be improved, and relevant

The national level should aim to build a safe AI industry chain and do a good job in individual

The development of basic standards such as personal data protection, cybersecurity, and AI ethics, and continued promotion of smart city human resources

Carry out standardization of national standards for intelligent industrial safety and improve the national standardization level.

8.1.2. Promote technological development and strengthen independent control

The development of smart cities is supported by a number of artificial intelligence technologies, from the entire life cycle of data to

From long-term management to information system construction, a large number of emerging technologies are introduced in every link.

From a different perspective, AI enables network security, data security, and application security, ensuring the security of various business systems in the city.

Allow different services to run smoothly in smart cities, making cities more efficient.

At the same time, with the rapid development of AI technology, in addition to traditional network security technology, AI capabilities

The security technology of the platform also needs to be developed urgently. For AI large model technology, such as adversarial training, model verification, The comprehensive application of security technologies such as design and verification, model encryption, and data masking can effectively

Improve the safety of the technology.

In the face of the rapidly changing artificial intelligence technology, my country should increase investment in core technologies and encourage scientific research institutions to

Institutions and enterprises conduct in-depth research on key technologies and products that are independently controllable, and promote domestic substitution

Degrees.

8.2. **AI** makes cities safer and smarter

8.2.1. Strengthen operational management and cultivate a team

The construction of smart city artificial intelligence involves a lot of software such as computing power, data, and servers.

Hardware assets, compared with simple data management, AI management is more complex and requires maximum

To maximize the value of these assets, the key is operation.

Establish a comprehensive city management system to integrate data from various city departments and implement data security protection

Second, promote intelligent transportation systems and smart

Security monitoring improves urban traffic management and public safety.

Train city managers and practitioners to improve their understanding and application of artificial intelligence technology

It is important to emphasize that the operation and management of AI in smart cities requires

It is necessary to combine it with social governance, build a smart city governance system, and promote information sharing and cross-departmental collaboration.

Achieve sustainable development of smart cities.

8.2.2. Improve **AI** security system and governance

We will continue to work hard to improve the AI-driven smart city security system and ensure that it becomes a company

part of daily operations to provide solid security support for AI infrastructure and business applications.

By clarifying the responsibilities and roles of all relevant parties, we will build comprehensive AI security governance and strive to meet the

Security goals include legality, fairness, trustworthy data security, and controllable and manageable systems.