Thinking Ahead About AI Security and Privacy Protection

Protecting Personal Data & Advancing Technology Capabilities

Huawei GSPO Office

Shield Lab, Huawei 2012 Laboratories

September 2019

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

CONTENTS

Executive Summary

Artificial intelligence (AI) has the ability to profoundly change every industry and every organization. However, while bringing substantial opportunities and benefits, this new general-purpose technology also faces significant challenges in security and privacy protection. The healthy development of AI relies on the governance of AI security and privacy protection. Huawei understands this and has been providing innovative ICT infrastructure and smart devices to carriers, enterprises, governments, and individual consumers worldwide over the past few decades, effectively promoting digital transformation and creating enriched value for society.

AI Definition and Scope

The industry has yet to propose a unified definition of AI. Basic characteristics of AI include handling complex goals, collecting and combining different amounts of data, extracting information and knowledge, learning autonomously, and making automated decisions at different levels. Commercial-level AI mainly involves two types of applications: internal enterprise applications used to improve work efficiency, and applications enabling vertical industries (for example, to improve automation, enhance capabilities, and inspire innovation). AI activities involve four key roles, namely, consumers/customers, deployers, solution providers, and data collectors. These characteristics, applications, and AI activities provide a basis for our governance of AI security and privacy protection.

Challenges

Industry standards and specifications indicate that technical reliability, societal applications, and legal requirements and responsibilities are the main challenges to AI security and privacy protection.

Huawei's Full-Stack, All-Scenario Solution

Huawei's full-stack, all-scenario solution (full-stack solution for short) introduces the AI mindset and technology into existing products and services. For example, AI is applied to the SoftCOM to improve O&M efficiency for carriers, enterprise intelligence (EI) provides a full-stack solution for enterprises and governments, and HiAI provides a full-stack solution for smart devices.

What We Did

Huawei has proposed a number of feasible governance practices, including planning trustworthy technical solutions, to address the possible challenges facing AI security and privacy protection.

Clarifying AI Stakeholders' Responsibilities for Building a Digital World

We will channel our expertise into bolstering AI security and privacy protection. This is just the beginning.

We call on governments, standards organizations, end users, and the industry as a whole to reach a consensus and work together to develop new codes of conduct, standards, and laws specific to AI and its use cases. Huawei and its global partners will work together to review their tasks based on business scenarios, further clarify responsibilities and activities, and provide systematic reasoning and governance methods to jointly provide people with AI services that can ensure security and privacy.

01 AI Definition and Scope

***Ethics Guidelines for Trustworthy AI*[1] released by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) in 2019:** Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

***Artificial Intelligence Security White Paper*[2] released by the China Academy of Information and Communications Technology (CAICT) in 2018:** AI enables intelligent machines or intelligent systems on machines. It studies and develops theories, methods, and technologies for simulating, extending, and expanding human intelligence, perceiving the environment, obtaining knowledge, and using knowledge to reach optimal results.

***Ethically Aligned Design First Edition*[3] released by IEEE in 2019:** ... the design, development, deployment, decommissioning, and adoption of autonomous or intelligent software when installed into other software and/or hardware systems that are able to exercise independent reasoning, decision-making, intention forming, and motivating skills according to self-defined principles.

***A Proposed Model AI Governance Framework*[4] released by Singapore Personal**

[1] Europe Commission's AI HLEG, *Ethics Guidelines for Trustworthy AI*, 2019.

[2] CAICT, *Artificial Intelligence Security White Paper*, 2018.

[3] IEEE Standards Association, *Ethically Aligned Design First Edition*, 2019.

[4] Singapore PDPC, *A Proposed Model AI Governance Framework*, 2019.

Data Protection Committee (PDPC) in 2019: "Artificial Intelligence (AI)" refers to a set of technologies that seek to simulate human traits such as knowledge, reasoning, problem solving, perception, learning and planning.

What is AI: No unified definition yet. Basic characteristics of AI:

(1) Handling complex goals

(2) Collecting and combining different amounts of data

(3) Extracting information and knowledge and learning autonomously

(4) Making automated decisions at different levels

AI is transforming numerous industries and having a profound effect on them. Although AI is currently not defined uniformly, it is being expanded and industrialized based on its common characteristics. Examples of identified commercial-level AI applications are as follows:

(1) Internal enterprise applications — improving work efficiency:

- Customer service: AI integrates with automated hotlines that offer a 24/7 communication channel for customers. AI classifies customers, integrates customer data, and manages activities.

- HR: AI assists in the recruitment process to match the supply and demand of candidates, screen candidates' resumes, and intelligently recommend potential candidates. Internal management can also enhance career development plans for employees by establishing competency & qualification (C&Q) profiles and an evaluation tracking system, providing a reference for internal transfer and promotion. AI not only increases the work efficiency of HR, but also improves the work experience.

- IT service monitoring: AI can automatically detect system exceptions and determine the optimal solution if IT faults occur, improving the work efficiency of O&M personnel.

- Supply chain management: AI supports predictive resource allocation and scheduling, proactively adjusting inventories to handle issues in the supply chain and sending reminders to supply chain planners.

(2) Applications enabling vertical industries — improving automation, enhancing capabilities, and inspiring innovation:

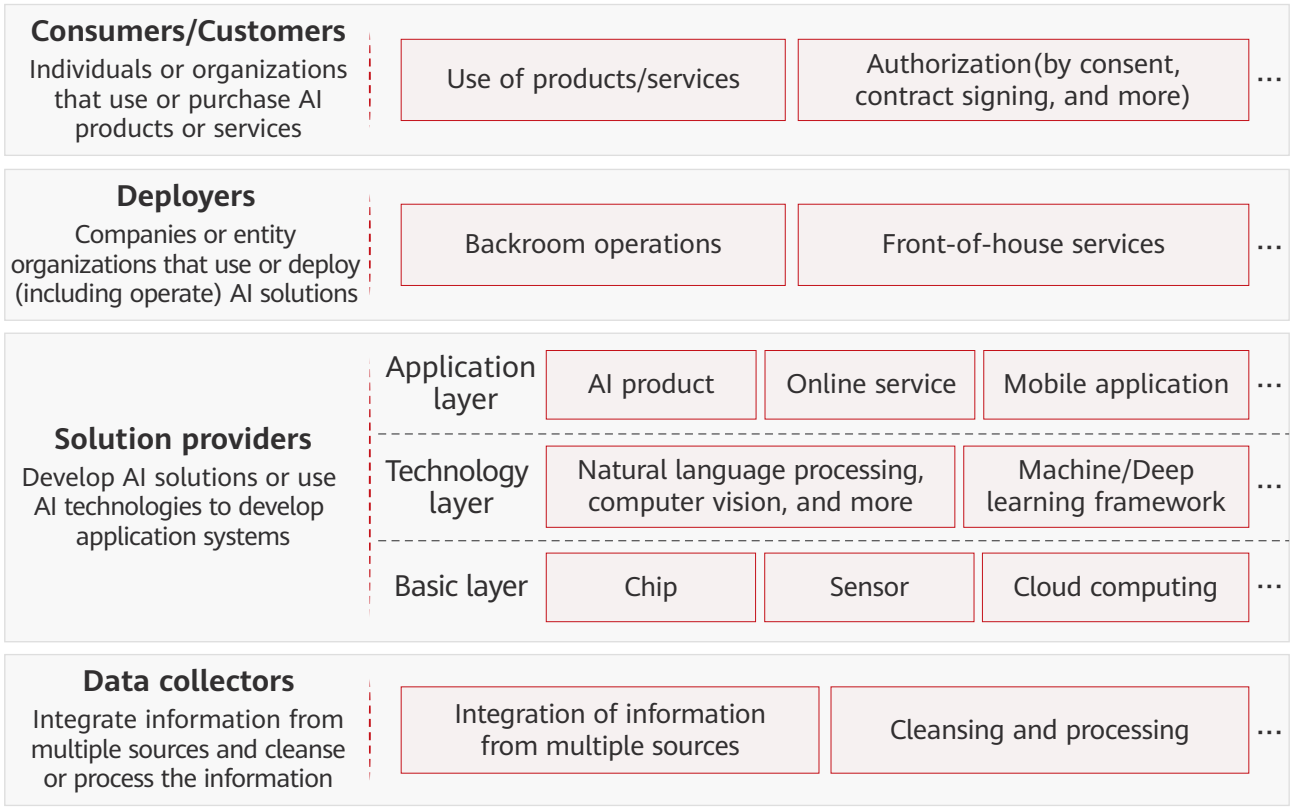
- Healthcare: AI can be used in medical image recognition, diagnosis using medical images, disease prediction, and risk analysis.
- Retail: AI optimizes the configuration efficiency of industry chain resources for production, process, and sales via intelligent customer service, intelligent payment systems, unstaffed warehousing, and more.
- Transportation: AI helps autonomous vehicles establish an accurate location, generate real-time models of the vehicle surroundings, and plan driving routes and strategies by sensing the external environment and reading engine data.
- Manufacturing: AI improves the productivity of production lines and helps restructure production capacities. By using computer vision and machine learning technologies for mechanical testing, AI reduces machine downtime and consequently lowers the OPEX.
- Smart City: AI can be used to collect city management data, including information about the atmosphere, water quality, lighting, transportation, schools, communities, and hospitals. This data provides municipal management personnel with valuable insights into city conditions, enabling them to plan and schedule resources more effectively.

How AI is used:

- Improving enterprises' efficiency
- Enabling vertical industries

The roles, activities, and scope of AI are summarized as follows.

Reference architecture of AI roles, activities, and scope (for commercial use)



Solution providers cover three layers, where they perform different tasks:

- **Basic layer:** Provide computing power. This layer mainly includes chips, sensors, and cloud computing. Chips have a high technical threshold and are the main enabler of computing power.

- **Technology layer:** Develop application technologies, including speech recognition, natural language processing, computer vision, and machine learning, for different fields. This layer relies on the computing platform and data resources for massive recognition training and machine learning modeling.

- **Application layer:** Solve practical issues — AI provides targeted products, services, and solutions — with commercialization as the core.

Four key roles in AI activities:

◆ Consumers/Customers ◆ Deployers ◆ Solution providers ◆ Data collectors

02 Challenges to AI Security and Privacy Protection

Technical reliability, societal applications, and legal requirements and responsibilities are three broad challenges facing the security and privacy protection of AI development. These broad challenges involve a number of smaller challenges, which are described as follows:

Three challenges facing AI development:

- Technical reliability
- Societal applications
- Legal requirements and responsibilities

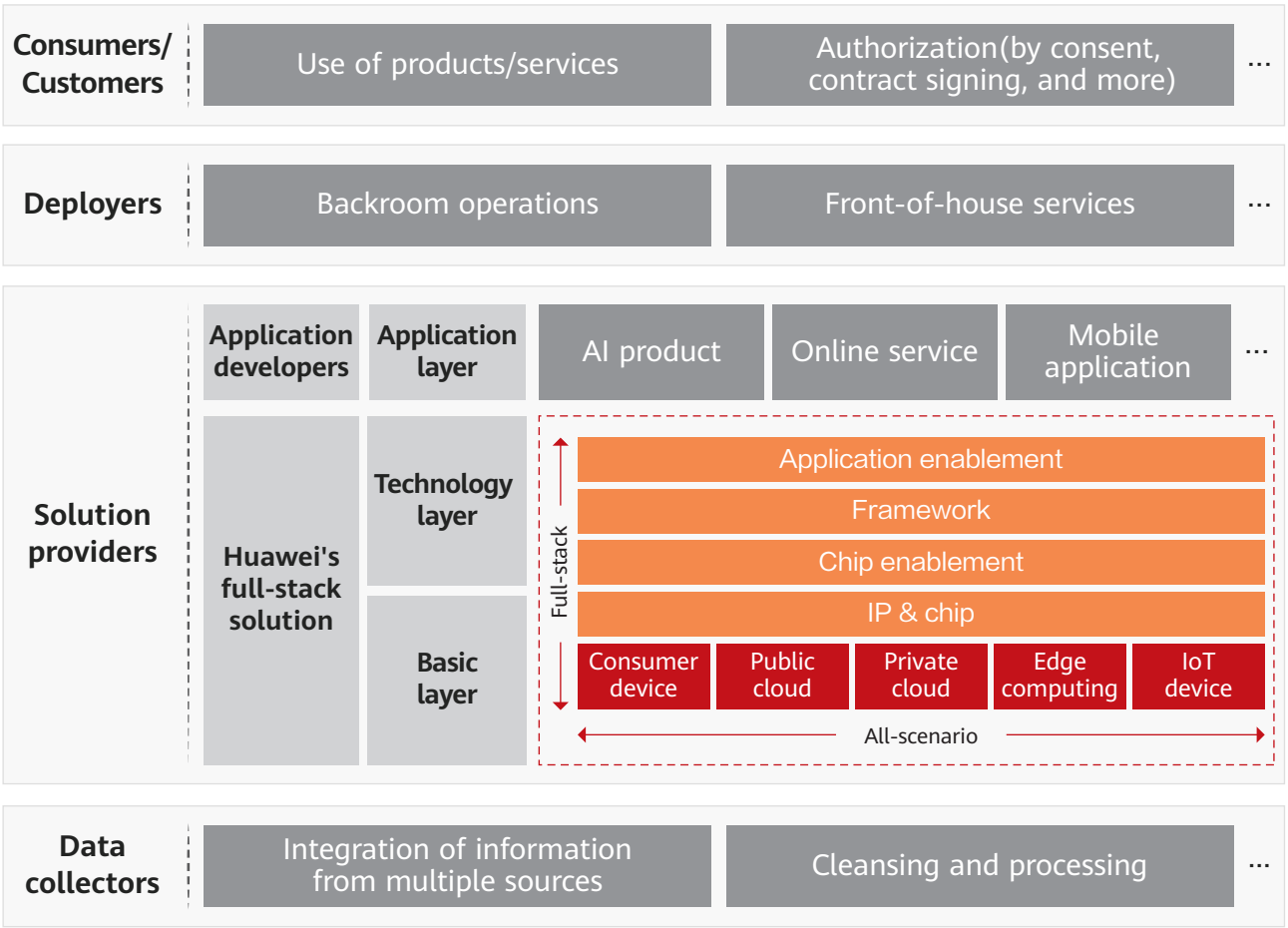
	Challenges	Cases
Technical reliability	• Deep neural networks (DNNs) lack robustness and may therefore be susceptible to evasion attacks. Such attacks will impair the judgment of AI systems and affect business security. • Complex systems such as DNNs inherently lack transparency and explainability, which may infringe upon legal or regulatory requirements (such as GDPR in terms of automated decision-making) and may even cause potential unfairness, inaccurate or unidentifiable results, and untraceable or unaccountable consequences. • Huge volumes of data may be originated from diversified operating environments. Data breaches, tampering, theft, and misuse may result from the unavailability of comprehensive data security protection.	• In the field of autonomous driving, evasion attacks can lead to traffic offenses and even accidents. • Attackers can introduce significant errors in the dosage recommended by AI models for nearly 50% of patients by adding only a small amount of malicious data. • Although the accuracy rate of cancer screening by AI may be high, doctors agree with about only half of the results because they think the results lack reasoning and logic.

	Challenges	Cases
Societal applications	• The lack of management and control over the purposes of AI may lead to AI being misused. • Data quality issues may lead to biased and unfair judgments. • Application developers and deployers who have insufficient knowledge and capabilities may misuse AI systems or cause security and privacy incidents.	• AI helps people who have difficulty in making sounds create a realistic voice. However, fraudsters may record a person's voice and use AI to generate speech for the purpose of phone fraud. • Credit scores are determined by collecting and analyzing a person's network behavior. Immigrants who are not proficient in English were assigned poor credit scores. • Facial recognition software incorrectly matched the photos of government officials with criminals' mug shots with a high false match rate due to improper parameter settings.
Legal requirements and responsibilities	• No laws or regulations, such as regulations on autonomous driving and algorithm accountability, are available to clearly define the rights and responsibilities of stakeholders.	• The AI algorithm used to establish a patient's drug dosage can infer the genetic information of the patient according to the drug dosage, breaching the patient's privacy. • An autonomous driving vehicle killed a pedestrian, leading to heated discussions about the supervision and legal responsibilities of autonomous driving vehicles.

03 Huawei's Full-Stack Solution and Applications

According to the reference architecture of AI roles, activities, and scope (for commercial use), the scope and activities involved in Huawei's full-stack solution are illustrated as follows.

Reference architecture of AI roles, activities, and scope



The full-stack solution covers four layers: Ascend, Compute Architecture for Neural Networks (CANN), MindSpore, and application enablement. Ascend — the basic layer of the full-stack solution — is the IP and chip layer. It aims to provide optimal performance at minimal cost for all scenarios. The CANN layer offers a chip operator library and highly automated operator development tools. It aims to provide optimal development

efficiency and operator performance to handle the rapid development of academic research and industry applications. The MindSpore layer is a unified training and reasoning framework that supports device, edge, and cloud (both standalone and collaborative). It aims to be design-friendly, operations-friendly, and applicable to all scenarios. The application enablement layer is a machine-learning platform as a service (PaaS) that provides ModelArts services, hierarchical application programming interfaces (APIs), and pre-integrated solutions. It aims to meet the needs of different developers by simplifying the aspects related to AI.

Over the past year, Huawei made a number of breakthroughs in relation to the full-stack solution: Released an AI model market based on ModelArts that provides industry-leading performance; released MindSpore for beta test; optimized CANN to support hundreds of operators and mainstream frameworks such as TensorFlow; launched the Ascend 910 chip, Ascend-Lite-based Kirin 810 chip, and products built on the Ascend 310 chip and related online services; implemented full-stack AI in more than 200 projects across over 10 industries.

Objective of the full-stack solution:

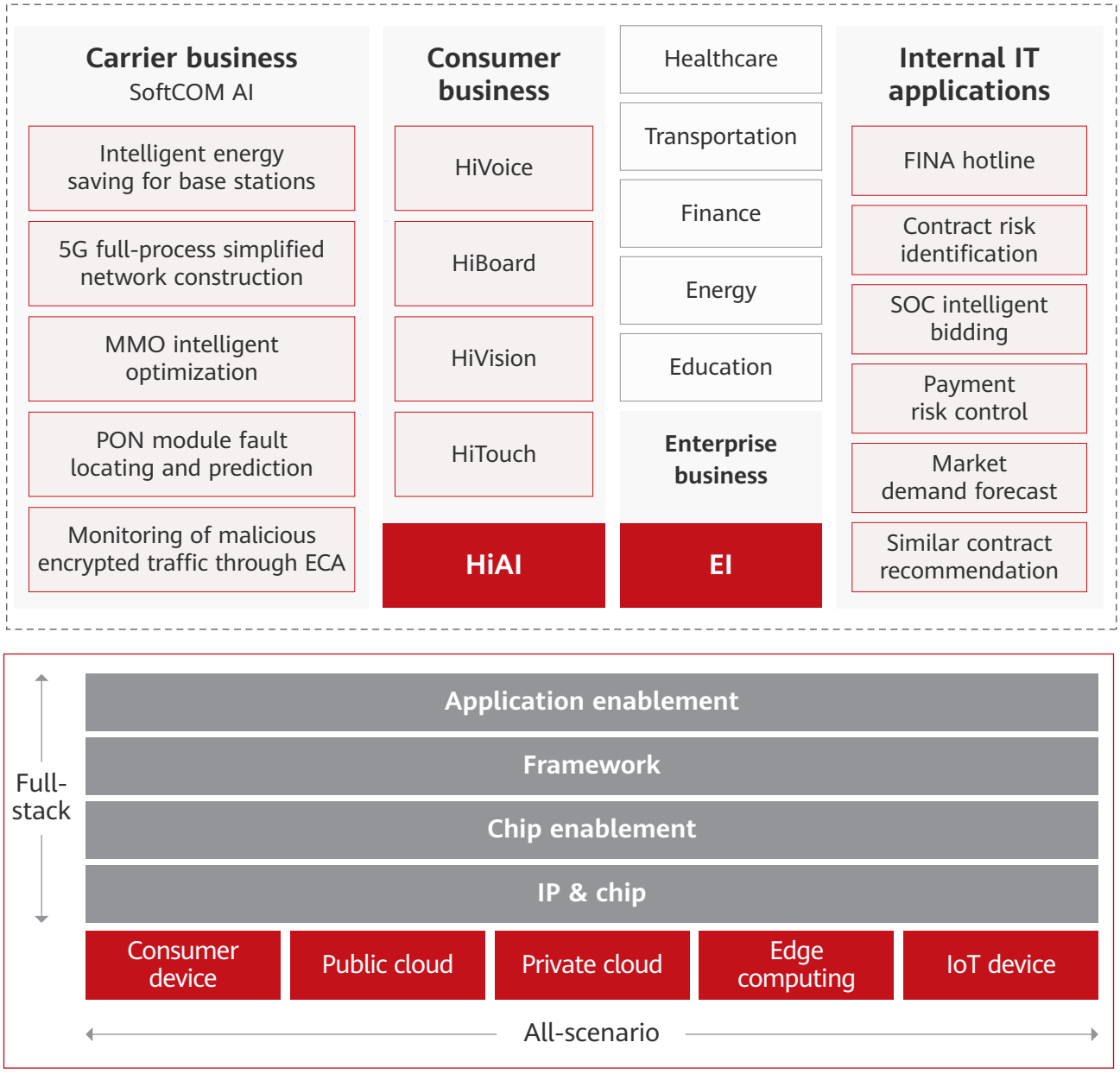
Provide affordable, effective, and reliable AI to realize inclusive AI

Huawei introduces the AI mindset and technology into existing products and services. For example, AI is applied to the SoftCOM solution to improve O&M efficiency for carriers, EI provides a full-stack AI solution for enterprises and governments, and HiAI provides a full-stack solution for smart devices.

Huawei's AI strategy:

- Invest in AI research
- Build a full-stack solution
- Develop an open ecosystem and talent
- Enhance the existing solution
- Improve O&M efficiency

Huawei's full-stack, all-scenario AI solution and applications



AI-driven E2E simplification of 5G network deployment is used to illustrate the Huawei SoftCOM solution. AI is applied throughout the entire 5G network deployment process — from site planning and deployment, to network optimization and maintenance — greatly simplifying the process. During site planning, the coverage status of base stations and the distribution of sites across the entire network are visualized and measurable. During site deployment, hardware is automatically discovered and configured, cells are automatically bound, and radio parameters are

automatically configured, implementing automated 5G site deployment. During network optimization, networks accurately identify the different coverage scenarios and automatically adjust relevant parameters to achieve optimal network performance based on the optimization objective, 3D map, online map, and 5G traffic map. During network maintenance, the system automatically analyzes network logs and alarms to identify and predict network faults. AI-based alarm automation minimizes the mean time to repair (MTTR) and field service. In addition, the time required to perform root cause analysis (RCA) is shortened by 80%, power consumption is reduced by 10% to 20%."

SoftCOM for carriers:

Build intelligent self-healing and self-optimizing networks to achieve network autonomy

HiAI is an open AI capability platform for smart devices. Adopting the "chip-device-cloud" architecture, it opens up chip, application, and service capabilities for a fully open intelligent ecosystem. This enables developers to quickly leverage the powerful AI processing capabilities on offer from Huawei. Kirin 980, Huawei's powerful dual-core neural-network processing unit (NPU), empowers the HiAI platform to recognize 4500 images per minute, doubling the original image recognition capability. With an increased number of supported development models and operators, AI applications can run faster on mobile phones and deliver a better user experience. The HiAI platform empowers Huawei to provide diversified AI products and services for consumers, facilitating activities both at home and at work and improving user experience. For example, HiVision is an image recognition app that provides a number of useful services, such as enabling users to scan barcodes, QR codes, and even objects, products, and food to identify them (even the number of calories contained in the food).

HiAI for consumers:

An open AI capability platform for smart devices, enabling developers to deliver a better experience to users of smart applications by quickly leveraging Huawei's powerful AI processing capabilities

EI is the enabler of enterprise intelligence. It provides an open, trustworthy, intelligent platform based on AI and big data through cloud services (such as public cloud and private cloud). EI, combined with industrial scenarios, enables enterprise application systems to view, hear, speak, analyze, and understand pictures, videos, languages, texts, and more. This allows more enterprises to experience the convenience of AI and big data services and ultimately accelerates business development for the benefit of society. For example, the HUAWEI CLOUD Autonomous Driving Cloud Service (Octopus) provides autonomous driving data, models, training, simulation, and data labeling services that help speed up the development of autonomous driving products.

EI for enterprises:

Use cloud services to provide an open and trustworthy AI platform, facilitating the use of AI by more enterprises and enabling EI

Huawei uses AI technology to improve its internal operational efficiency in repeated, massive, and complex IT business scenarios. We leverage AI in HR, finance, supply chain, and other scenarios across multiple aspects and levels. For example, HR intelligent customer service improves user experience, and machine learning establishes strategic rules for monitoring risks and automating how vast numbers of delivery operations are processed.

Internal IT applications:

Use AI technology to improve Huawei's internal operational efficiency

Huawei also uses AI technology to enhance the performance of its security products and improve the advanced threat detection accuracy to more than 99% for the original AI threat detection engine. The in-house AI chip improves computing power, increasing the advanced threat detection efficiency by five times. Distributed AI joint detection is continuously updated and optimized, clearing local unknown security threats in real time. In addition, intelligent security analysis and defense reduce the OPEX for secure

operations and maintenance by 80%.

Enhancement of security and privacy protection capabilities:

Use AI technology to enhance the performance and competitiveness of Huawei's security products

04 Huawei AI Security and Privacy Protection Governance Practices

Although AI brings substantial opportunities and benefits, we need to develop feasible governance that can address security and privacy challenges.

The overall AI governance objectives are mainly considered from seven aspects:

System security and controllability: The system meets the security requirements of robustness, stability, and adaptability and can provide security attestation.

Transparency and traceability: Logics (such as prediction, judgment, and automated actions) generated by the system must be explainable and transparent. During operations, data and activities that involve security and privacy must be documented based on application scenarios to ensure that the judgment logic and action process are demonstrable and accountable.

Privacy protection: Personal information and data are protected and managed in compliance with the GDPR and other applicable laws. Data and information are shared and AI development is accelerated while privacy is ensured.

Fairness: Unfairness is inherent in society or caused by procedures. To mitigate unfairness inherent in society, ensure fair distribution of benefits and costs, and avoid bias, discrimination, and insult. To mitigate unfairness caused by procedures, challenge decisions made by AI and seek effective remedies.

Data management: Data is critical to AI judgment results and model generation. Ensure the integrity, accuracy, availability, confidentiality, and comprehensiveness of data.

Competence: A solution provider should specify the knowledge and skills required to securely and effectively run AI, and those who deploy and develop solutions should follow these knowledge and skills.

Deployment objective assurance: AI deployment and application must meet the requirements of lawfulness, fairness, security, privacy protection, and no misuse, as well as deployment purposes.

All AI stakeholders need to work together to achieve the aforementioned objectives. Huawei has made ongoing efforts to fulfill the security and privacy protection governance responsibilities regarding the full-stack solution, and will continue to improve and supplement these practices.

Overall governance objectives:

- System security and controllability
- Transparency and traceability
- Privacy protection
- Fairness
- Data management
- Competence
- Deployment objective assurance

Huawei AI security and privacy protection governance practices

Security and privacy protection policies	Organization and personnel	Process and product design	Technology	Verification
• Strategy • General cyber security policy • General privacy protection policy • Data sharing and control policy	• Organization settings • Certifications related to the full-stack solution • Awareness and capabilities	• Risk analysis • Requirements built into processes and products • Design specifications	• Application security enablement • Trustworthy framework • Trustworthy operator • Trustworthy chip	• Data sharing and protection measures • Awareness and capabilities of full-stack solution users • Explainability and traceability of the full-stack solution

Cyber security and user privacy protection are and will remain Huawei's highest priorities.

—— Mr. Ren Zhengfei, Founder and CEO of Huawei

1. Strategy: We will always prioritize security and privacy over costs, schedules and functions. We imbue security processes into product lifecycles: from design, to development, to delivery. Huawei is committed to invest US$2 billion over the next five years to build secure, trustworthy, and high-quality products in our ICT infrastructure business. Huawei has signed cyber security agreements with more than 3400 cyber security-related suppliers worldwide — requiring all partners to take security seriously — and has also signed data processing agreements (DPAs) with more than 1300 suppliers.

2. General cyber security policy: We've fully committed to security in every way. We welcome input, ideas and suggestions to improve everything we do, so we can benefit our customers, and their customers. Today, we're probably the most open, most evaluated and transparent company in the world. We will always prioritize security over costs, schedules and functions. We imbue security processes into product lifecycles: from design, to development, to delivery. If there is a security standard or security certification that needs to be achieved, we will achieve it. We believe you cannot have good privacy without good security, nor good security without good privacy. We would never do anything illegal. We will never harm any country or any individual, and never accept any request to use Huawei products for malicious purposes. If we are ever put in such a position, we would rather close the business. We're here to help our customers maximize the value of their assets. Nothing matters more to us than being customer-centric. It's why we do more to build trust, to enhance our capabilities, to be transparent, and advocate collaboration. Security isn't just something we invest in constantly, but a value that serves as the foundation of our existence.

3. General privacy protection policy: Huawei's privacy protection framework sets differentiated privacy protection objectives based on the characteristics of different

business domains. We incorporate privacy protection requirements into the E2E IPD process and require all product versions to strictly comply with specified regulations. Our Privacy Protection Guideline is updated yearly (we have released four versions to date).

High-risk businesses apply a "dual-officer approval" mechanism to their privacy protection decision-making. For example, HiAI must be approved by the Global Cyber Security & Privacy Officer (GSPO) and Chief Legal Officer (CLO) before it is deployed.

Huawei adheres to the principle of "local data processing" by implementing the GDPR's data transfer agreement (DTA) framework during cross-border transfers of personal data.

4. Data sharing and control policy: The global cross-department data sharing and data classification management policy includes the classification criteria, definitions of each category, and control requirements for each phase of the data lifecycle. The policy describes data owners' and users' information security responsibilities, defines basic rules and exceptions for data sharing, and establishes pop-up arbitration and operations mechanisms for data sharing, facilitating data sharing in all data foundation areas. The areas covered in the policy also include the management and identification of personal data, the development of management capabilities for highly confidential and sensitive data, the labeling of personal privacy data before entering data lakes and during authorization, as well as operational capacity building (providing the data steward, tenant administrator, data foundation operations module by privacy professional, and more). In addition, the policy contains various detailed rules, such as centralized management rules for training data, data import compliance guidelines, and personal data retention specifications.

4.1 Security and privacy protection policies

Build and implement an E2E global cyber security and privacy protection system

1. Organization settings: The GSPO is an important member of the GSPC and directly reports to the CEO. The specific responsibilities of the GSPO are defines as follows:

leading the team in developing security and privacy strategies; planning, managing, and overseeing security and privacy organizations as well as services of R&D, supply chain, marketing and sales, engineering delivery, and technical service departments; ensuring the implementation of the cyber security and privacy protection system in various systems, regions, and processes; and actively promoting communication with governments, customers, partners, employees, and other stakeholders.

2. Full-stack solution related certification: To enable application developers to better understand the capabilities and system requirements of full-stack AI, Huawei provides targeted training and documented guidelines for its downstream players. For example, HCIP-AI EI Developer certification aims to cultivate professionals that can develop and innovate AI products by using Huawei enterprise AI solutions such as Huawei Cloud EI, general open-source frameworks, and Huawei's one-stop development platform for AI developers, ModelArts. In addition, HCIP-AI HiAI Developer certification aims to cultivate professionals that can use Huawei's device-chip-cloud three-layer open mobile computing platform HiAI for development and innovation.

3. Awareness and capabilities: Huawei released a privacy awareness video tutorial and examination, and about 98% of all employees passed the examination (170,000+). This examination will be optimized and held yearly, and currently more than 200 privacy professionals have obtained IAPP certifications, such as CIPP/E, CIPP/US, CIPM, and CIPT. In addition, Huawei provides focused methods and training to help developers understand AI explainability and the necessity of algorithm traceability.

4.2 Organization and personnel

The Global Cyber Security and User Privacy Protection Committee (GSPC) is Huawei's highest-level organization for managing cyber security and user privacy protection.

1. Risk analysis: It is difficult to manage AI risks because we are unaccustomed to them. Therefore, rather than creating new processes to deal with AI risks, we should

improve existing processes, incorporate AI factors into risk assessment, and develop workflows and handling tools to ensure that risks are effectively identified and managed within the overall framework scope.

In 2019, the European Commission's AI HLEG issued the Ethics *Guidelines for Trustworthy AI*[1]. Huawei selected the most relevant businesses for risk assessment and requirement identification.

2. Requirements built into processes and products: Huawei globally established standardized business processes and identified global process owners (GPOs) as well as key control points (KCPs) for each process. In addition, Huawei established a Global Process Control Manual and a Segregation of Duties Matrix that are applicable to all subsidiaries and business units. The GPOs are responsible for ensuring overall internal control effectiveness with regard to changes in operational environment and risk exposures. Control point examples include incorporating the performance indicator requirements of security and controllability into the R&D and design processes, and incorporating control points of privacy impact assessment (PIA) requirements into the processes. Registration and record storage systems should be created to record core parameters, helping identify persons responsible for specific AI. In addition, AI manufacturers, operators, and owners should register key high-level parameters, which include: intended use; training data and environment (if applicable); sensors and real world data sources; algorithms; optimization goals, loss function, and reward function; model features (at various levels); user interfaces; actuators and outputs; and process graphs. Furthermore, AI systems should generate audit trails that record facts and support decisions. These systems can be verified by third parties that use system logs and black-box recorders to confirm that the audit paths reflect what the system actually does.

[1] Europe Commission's AI HLEG, *Ethics guidelines for trustworthy AI*, 2019.

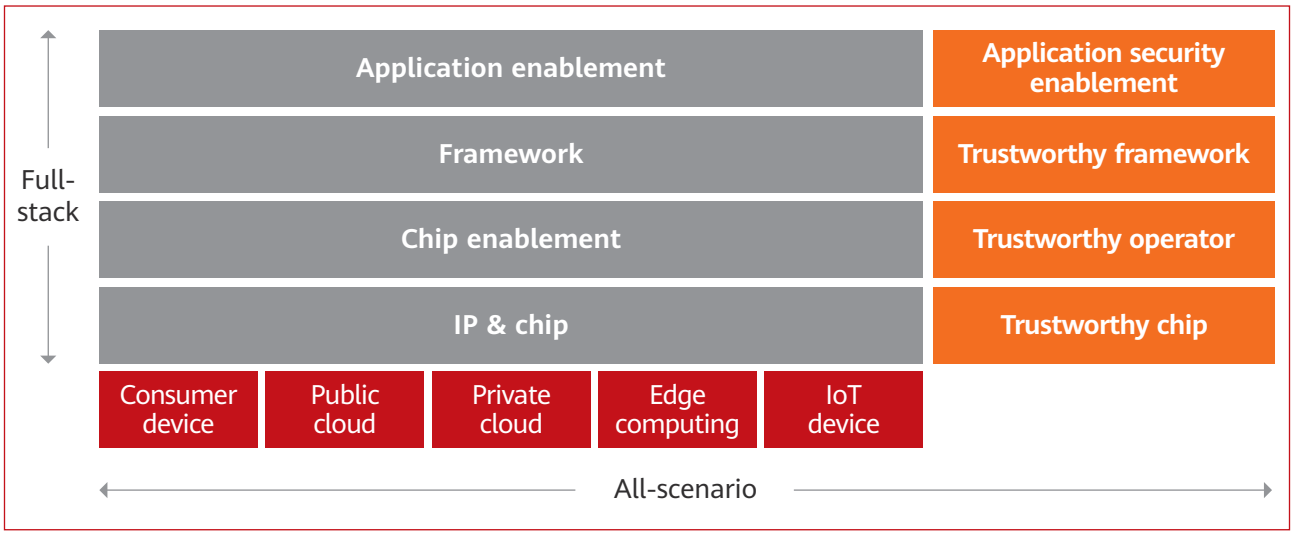
3. Design specifications: Ensure that products meet security and privacy protection requirements by integrating these requirements throughout the entire development process, from design to innovation. In addition, strengthen privacy protection technology innovation and maximize value creation under the premise of satisfying privacy protection. For example, leverage differential privacy technology to irreversibly anonymize application information so that application experience can be improved while ensuring the privacy of individual users.

Introduce AI-related design principles based on existing security design principles. After an AI system implements deep learning and training, various design-related issues may occur and AI might exhibit unintended behavior. Security and privacy risks, as well as implementation costs, will increase if these factors are only considered during the latter stages of design. Therefore, AI security and privacy protection must by default be built into each standard, protocol, and process of production and operations. Design principles include security design based on attack and defense, model security design, AI architecture security design, training and inference data privacy protection technology, and reliable AI system design and verification. For example, an AI system that autonomously makes decisions should be capable of secure continuous learning in dynamic environments. In such scenarios, continuously adjust and optimize the system during interactions based on feedback from each party, and set a multi-level security architecture to ensure overall system security. When the AI system's certainty is lower than the threshold, the system reverts to rule-based conventional technology to make decisions. A semi-autonomous AI system that makes decisions with human assistance must be equipped with a security recording module to record sensor data, user input, running status, and system output. This enables backtracking, diagnosis, and accountability when problems occur in the system.

4.3 Process and product design

Built-in security and privacy (instead of bolted on)

Security and privacy design objectives for Huawei full-stack solution: Data model integrity and confidentiality are protected by a secure runtime environment with hardware roots of trust; enablement of robust, traceable, and verifiable applications facilitates secure and reliable business deployment and operations. The corresponding four-layer requirements are as follows.



Application security enablement: Different business applications have varying security requirements, for example, secure model updates and data privacy protection. We need to provide comprehensive security capabilities and architecture for protecting business security, as well as a traceable and explainable solution that adapts to dynamic environments.

Trustworthy framework: The framework layer should provide privacy and security protection for developers and users. Security practices, such as security verification and maintenance, module function decoupling, and least privilege, are implemented to ensure that the framework is secure, transparent, and trustworthy.

Trustworthy operator: Models need defensive components and trustworthy algorithms (operators). A trustworthy chip operator library is used to provide security components and trustworthy operators for business logic, preventing attackers from stealing and tampering with model information through unauthorized call of operators.

Trustworthy chip: A trustworthy runtime environment with hardware roots of trust is essential to business security. We hope to build a secure and trustworthy runtime environment with Ascend series chips serving as universal and extensible hardware roots of trust, protecting the integrity and confidentiality of models and data.

4.4 Technology

Trustworthy full-stack solution

Huawei reviews and verifies AI systems and products from the following perspectives, conducts regular inspections, and makes appropriate improvements and adjustments:

- **Data sharing and protection measures:** Continuously monitor all data activities and take necessary control measures, log the operations of the entire data lifecycle on the data foundation and data analysis platform, and ensure the logs are integrated into the company's monitoring platform for centralized management.
- **Awareness and capabilities of full-stack solution users:** Promptly adjust and optimize AI application and control by regularly collecting user feedback on the full-stack solution, including opinions and expectation for AI capabilities, security and privacy issues, and trustworthiness of AI decisions.
- **Explainability and traceability of the full-stack solution:** Validate multiple system associations or provide valid and appropriate evidence for explainability and traceability.

4.5 Verification

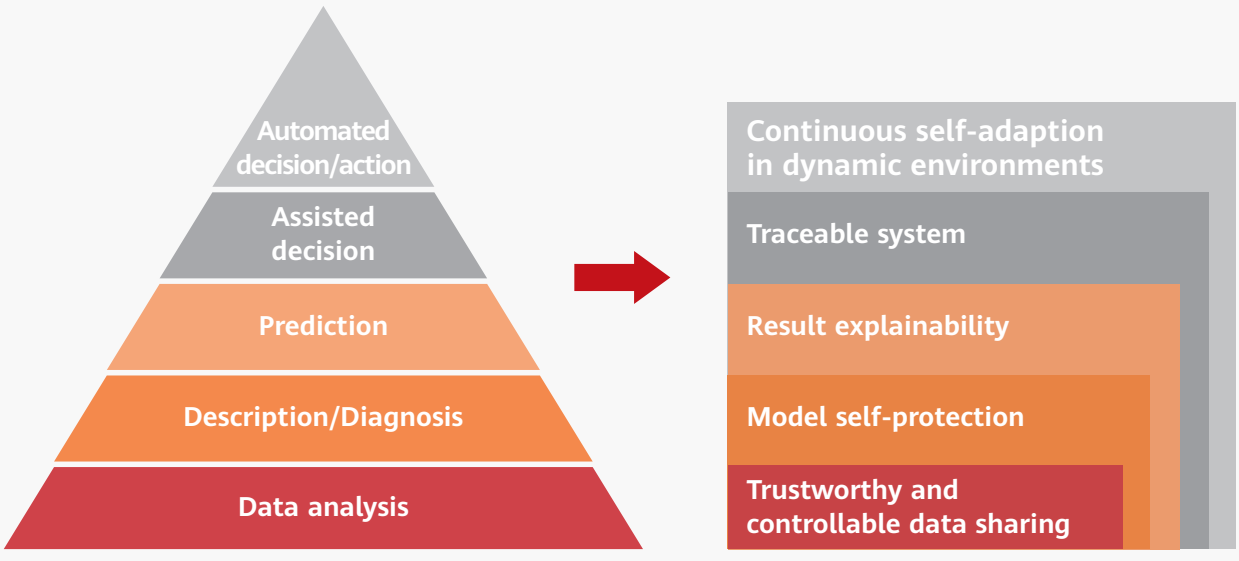
Assume nothing, believe no one, and check everything

05 Thoughts and Recommendations on Security and Privacy Protection for AI Application Development

AI capabilities are classified into five levels: data analysis, description/diagnosis, prediction, assisted decision, and automated decision/action. Higher capability levels correlate to higher security and privacy risks. Therefore, AI applications should match different capability levels that satisfy different security and privacy protection solution objectives; however, solutions objectives are downward inclusive, which means the objectives of each level must satisfy the objectives of lower levels.

AI application capabilities must be classified into different levels, and solutions must be adapted to those capability levels.

Hierarchical AI application capabilities and solution objectives



Hierarchical AI application capabilities

Security and privacy protection solution objectives

Continuous self-adaption in dynamic environments

- Security reinforcement learning: Reduce the impact of adversarial examples by training multiple algorithms and reinforcement learning in batches, and ensure that intelligent twins meet security restrictions during learning and running by embedding restrictive security requirements or other measures.
- Ensemble learning technique: Train multiple models and detect their self-consistency to enhance attack resistance and ensure model operation security.
- Secure fallback mechanism: Set a multi-level security architecture to ensure overall system security. For example, when an AI system's certainty is lower than the threshold, the system reverts to rule-based conventional technology to make decisions.

Traceable system

- Data security label: Data is labeled to identify and trace data tampering.
- Model signature tracking: The signature verification mechanism provides model source tracing.
- Model watermark: Watermarks are built into DNN models.

Result explainability

- Explainable model: This is also referred to as ante-hoc explainability. An explainable model with a simple training structure and optimal explainability can be built, or explainability can be integrated into a specific model, enabling the model to be capable of explainability.
- Post-hoc explainability: This indicates the explanation of a trained machine learning model through local approximation, sensitivity analysis, and more. Post-hoc explainability is classified into global explainability and local explainability based on goals and objects.

Model self-protection

- Defense against adversarial examples: Adversarial examples that are generated using various technologies are added to the defense system through retraining or other measures.
- Model theft prevention: Security isolation, theft detection, output scrambling, and more are utilized to prevent attackers from stealing model data.
- Model security detection: Certain methods are used to detect models from multiple dimensions (such as accuracy and missed detection rate) based on business scenarios.
- Secure enclaves/Trusted execution environment (TEE): A hardware-based running area with secure isolation is provided to protect data training and model parameters against attacks from malicious applications.

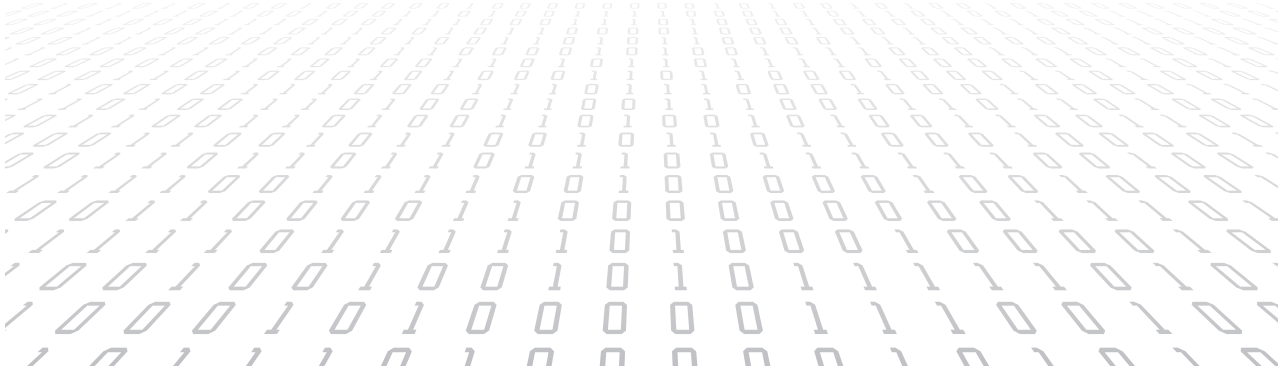
Trustworthy and controllable data sharing

- Cloud-edge collaboration: Data and models are shared within edge devices. Complex computing-intensive models are obtained through training results in the cloud, which ensures data and model security while making high-quality decisions.
- Differential privacy: Noise is added to ensure differential privacy (attackers cannot obtain any user's personal information based on results).
- Security federated learning: Security considerations are added to the federated learning framework to prevent malicious parties from obtaining additional information.
- Ciphertext learning: AI learns specially encrypted ciphertext to protect data security.
- Data filtering technology: Poisoned data is filtered out by analyzing the inherent nature of data.

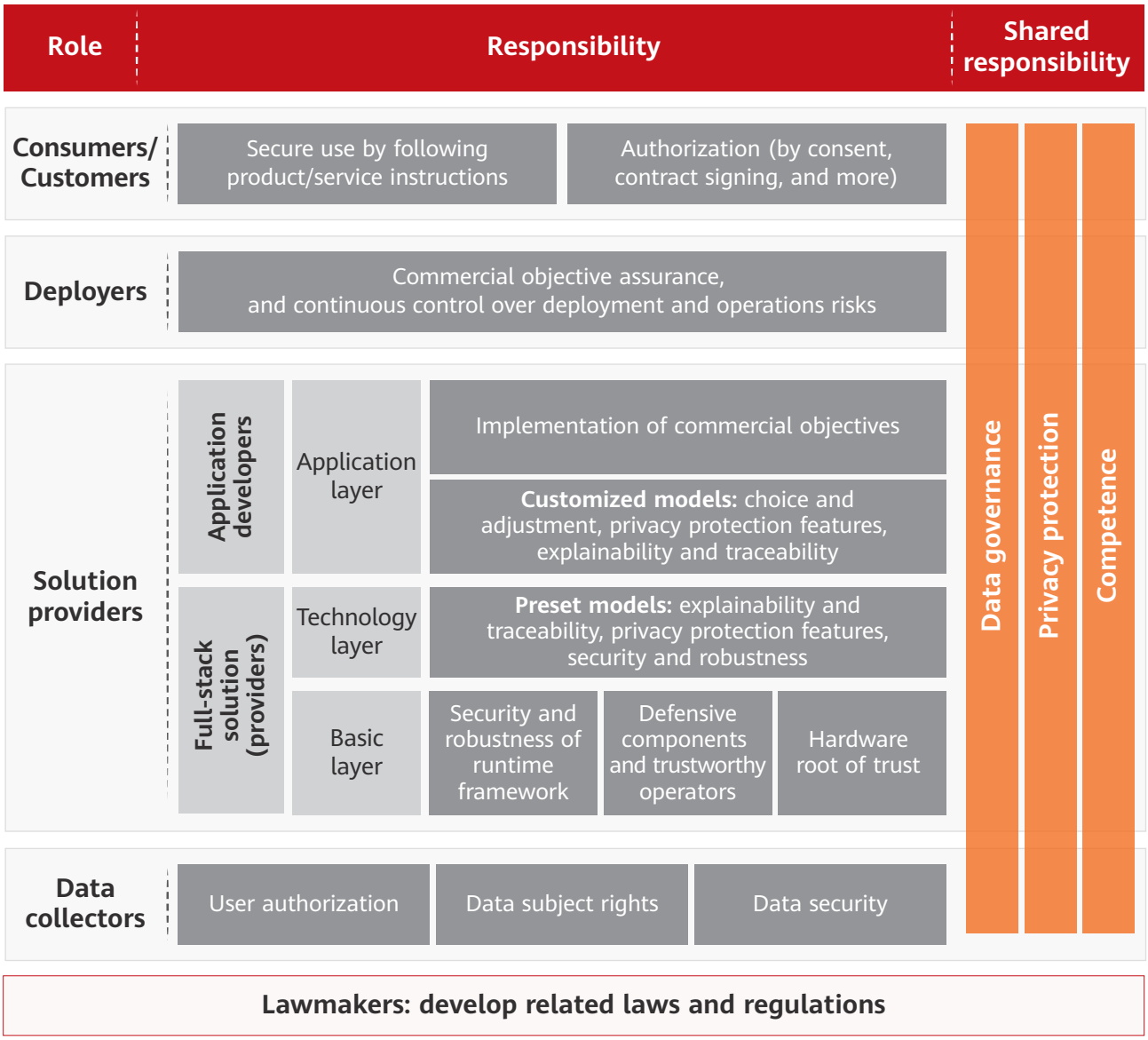
06 Clarifying AI Stakeholders' Responsibilities for Building a Digital World

The healthy and rapid development of AI can only be promoted through sharing responsibilities and clarifying the responsibilities of relevant stakeholders.

AI products and applications are part of a comprehensive ecosystem and encompass a wide range of market participants. Each market participant provides different ecosystem components and levels, which include tangible elements (such as chips, servers, and sensors), software, application operations and maintenance, as well as data processing and use. Therefore, every AI product or service is highly dependent on third-party technologies to fulfill its functions. The diversity of market participants may amplify the defects of each individual element and induce more issues caused by interaction between those elements. In addition, it is difficult for consumers and users to identify market participants' problems and corresponding root causes. Therefore, we call on global partners to review their own work based on business scenarios, further clarify their responsibilities and activities, and provide a systematic approach to thinking and governance based on shared responsibility. We recommend a shared responsibility model in which different activities are performed by different roles.



Shared responsibility model for AI security and privacy protection governance (recommended)



Relationships between governance objectives and responsible roles

To achieve overall security and privacy protection governance objectives, lawmakers, deployers, application developers, full-stack solution (providers), data collectors, and consumers/customers should govern their responsible activities based on their understanding of the objectives. The following is for reference purposes:

Lawmakers: Improve AI-related laws and standards, clarifying each stakeholder's responsibilities.

Deployers: Ensure deployment objectives as well as control security and privacy risks during deployment and operations.

Application developers: Provide critical support to deployers and the full-stack solution. Support application objective assurance and implementation for deployers; select algorithms and customized models based on application objectives and continuously adjust and update them for the full-stack solution.

Full-stack solution (providers): Assume responsibilities for the security and controllability of the solution. The technology layer mainly involves the explainability and traceability of preset models, privacy protection features, and security and robustness. The basic layer implements trustworthy solutions.

Data collectors: Guarantee compliance with the applicable laws and regulations of the data collection process (including obtaining user authorization and respecting data subject rights), data quality and security, and more.

Consumers/Customers: Assume responsibilities for the secure use of products or services in compliance with product or service instructions and authorization (by consent, contract signing, and more).

Shared responsibility: Privacy protection, data management, and competence are E2E implementation requirements that should be built into the business process and operations of each role.

07 Conclusion

From the definition of AI to the scope of Huawei's activities in AI, we aim to clearly define AI security and privacy governance from the source. From security and privacy protection governance strategies to trustworthy full-stack solution planning, we hope to promote the healthy development of AI, with Huawei leading by example. From the identification of overall governance objectives to the solutions that match hierarchical AI application capabilities, we aspire to do more for AI.

The purpose of the establishment, implementation, and improvement of security and privacy standards in the digital era is to achieve a harmonious coexistence between humans and nature while benefiting humankind. Huawei also recognizes that no single organization or company has sufficient resources to tackle increasingly complex security and privacy risks and threats to AI.

First, we call on governments, standards organizations, end users, and the industry as a whole to reach a consensus and work together to develop new targeted codes of conduct, standards, and laws to enhance AI security and privacy protection. Second, we hope that our partners can clearly define their responsibilities based on the concept of shared responsibility with Huawei and their business scenarios. We also hope that they can leverage their unique experiences and insights on AI security and privacy protection to improve the shared responsibility model with us.

We will channel our expertise into bolstering AI security and privacy protection. This is just the beginning. Huawei provides innovative ICT infrastructure and smart devices to global carriers, enterprises, governments, and individual consumers, effectively promoting digital transformation and creating enriched value for society. We also understand that it is not only vital to the healthy development of human society in the digital era, but also Huawei's incumbent responsibility to actively support the establishment, implementation, and improvement of security and privacy standards in the digital era, as well as help countries develop and establish their own digital era and processes.

HUAWEI TECHNOLOGIES CO., LTD.
Huawei Industrial Base
Bantian Longgang
Shenzhen 518129, P.R.China
Tel: +86 755 28780808

www.huawei.com