Standardization Profile and Engineering Practice of AI Governance

Liu Jiaqi, Ma Wanzhong

Institute of Strategic Research, Huawei

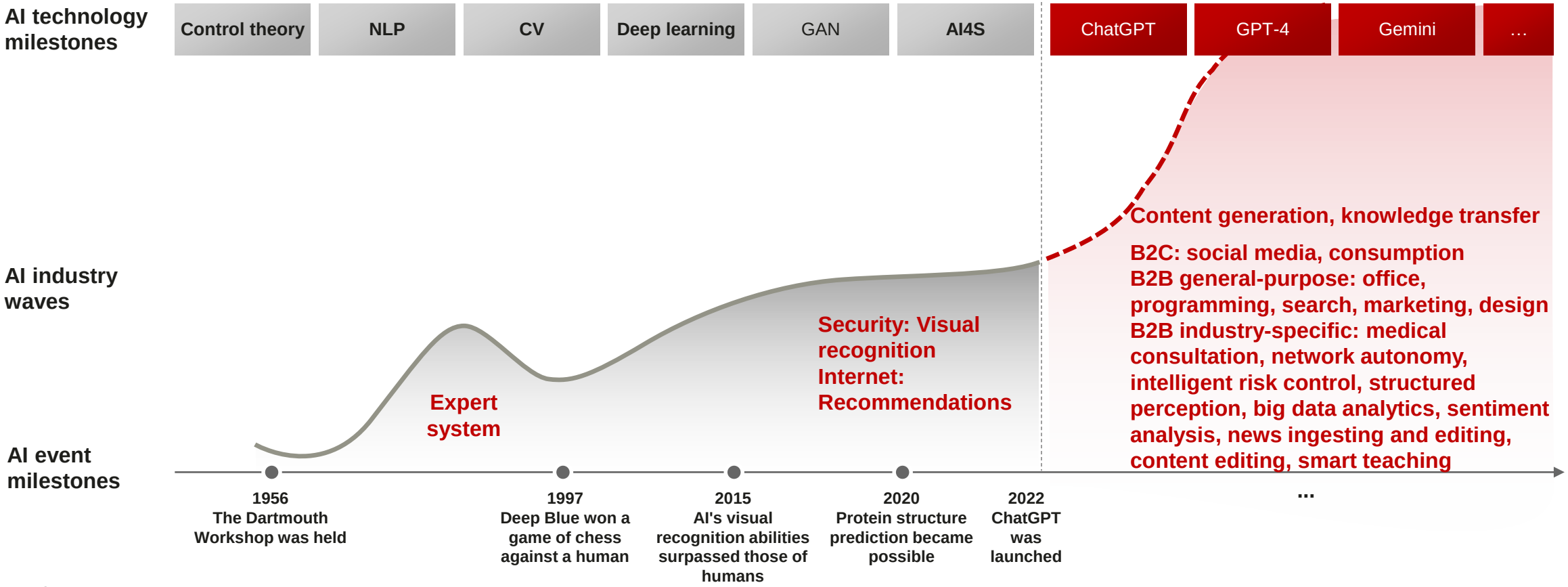
The evolution of AI: From prediction and inference to content generation, and from specific to general-purpose applications

Specific applications: Perception and understanding
Processing entry-level, repetitive work

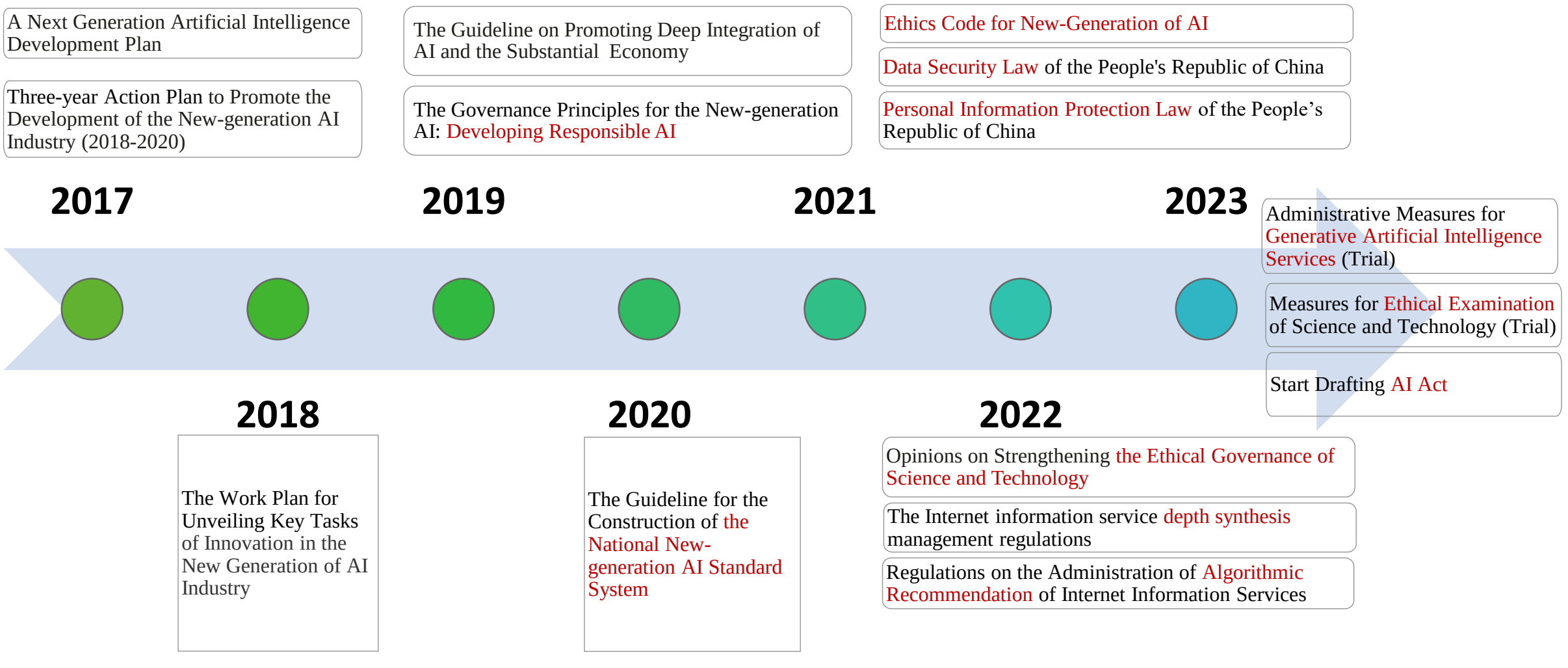
Partially replaces human eyes/ears; mainly used for object recognition and analysis; based on the convolutional neural network (CNN) architecture

General-purpose applications: Generation and creation
Processing complex, intellectual work

Capable of thinking and learning; mainly used for understanding and generating content; based on the transformer architecture



AI Governance Policy in China



AI standardization organizations in China

The Expert Advisory Group on AI National Standardization

- Consults on Chinese AI standardization plans, systems, policies and measures.
- Advises on the development, piloting and application of AI standards, and the establishment of standardization mechanisms for AI standards.
- Guides the work of the AI National Standardization General Group.

The AI National Standardization General Group

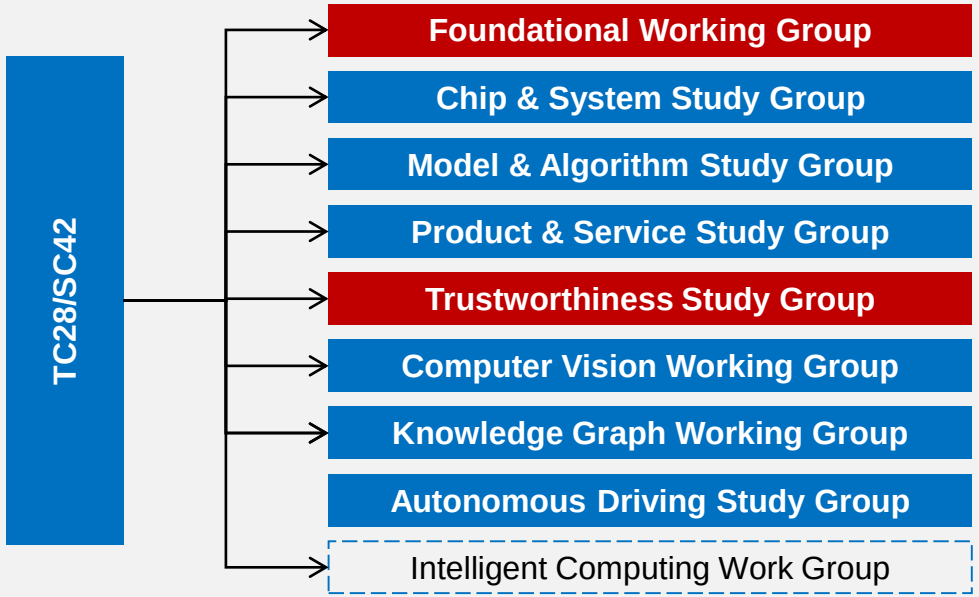
- Is responsible for formulating Chinese AI standardization plans, systems, policies and measures
- Coordinates the development of AI national standards between different industries and verticals
- Has their secretariat positioned at the China Electronics Standardization Institute (CESI)

The National Information Security Standardization Technical Committee

- In AI area, Is responsible for AI security related standardization
- Has their secretariat positioned at the China Electronics Standardization Institute (CESI)

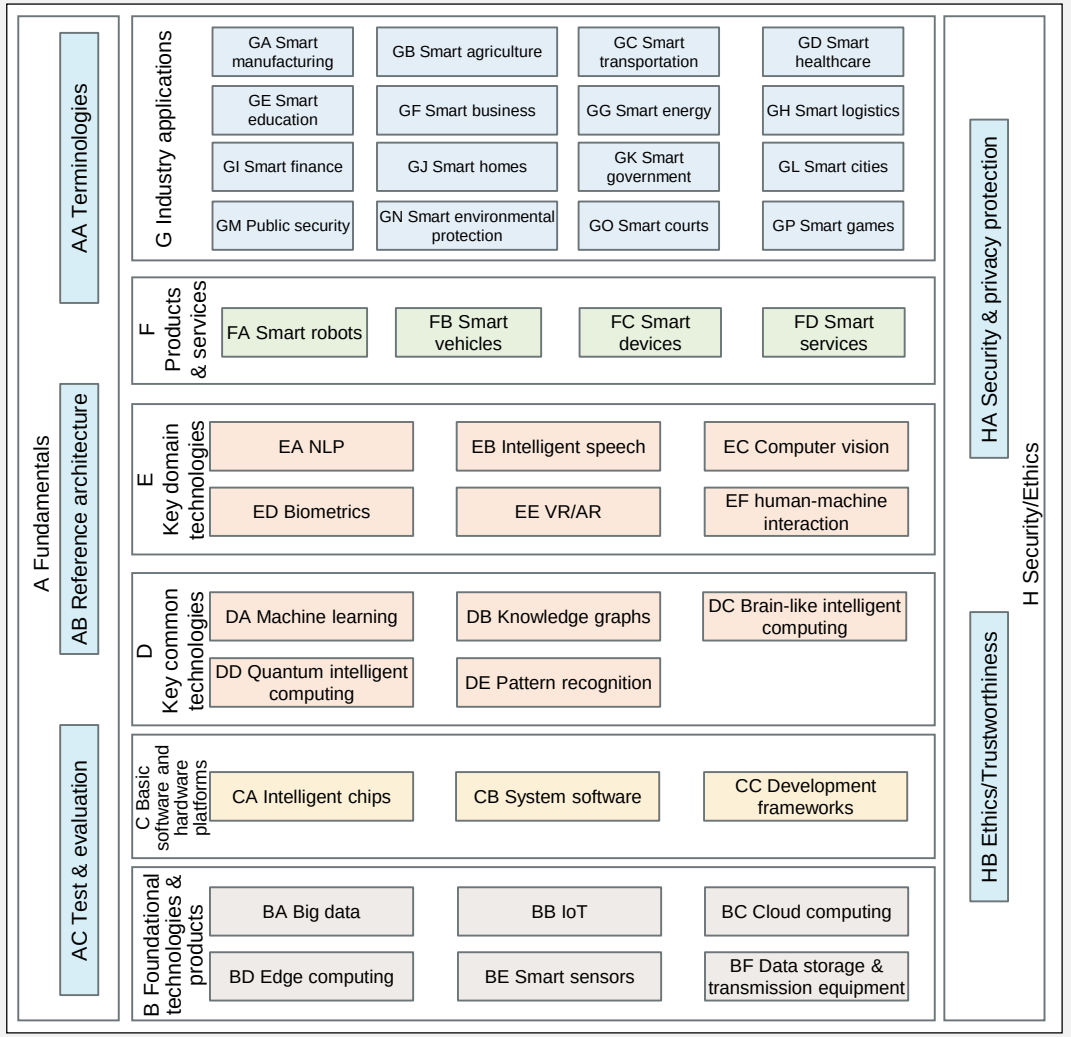
The National Information Standardization Technical Committee AI Sub Committee (TC28/SC42)

- Set up on March 18, 2020, with the approval of the Standardization Administration of China (SAC)
- Mainly responsible for the development and revision of national standards related to AI basics, technology, risk management, trustworthiness, governance, products, applications, etc.
- Corresponds to ISO/IEC JTC 1/SC 42 AI
- Has their secretariat positioned at the CESI



AI Standardization System and Architecture

- In July 2020, five departments, including the Standardization Administration of China (SAC) and the Ministry of Industry and Information Technology (MIIT), jointly issued the ***The Guideline for the Construction of the National New-generation AI Standard System.***

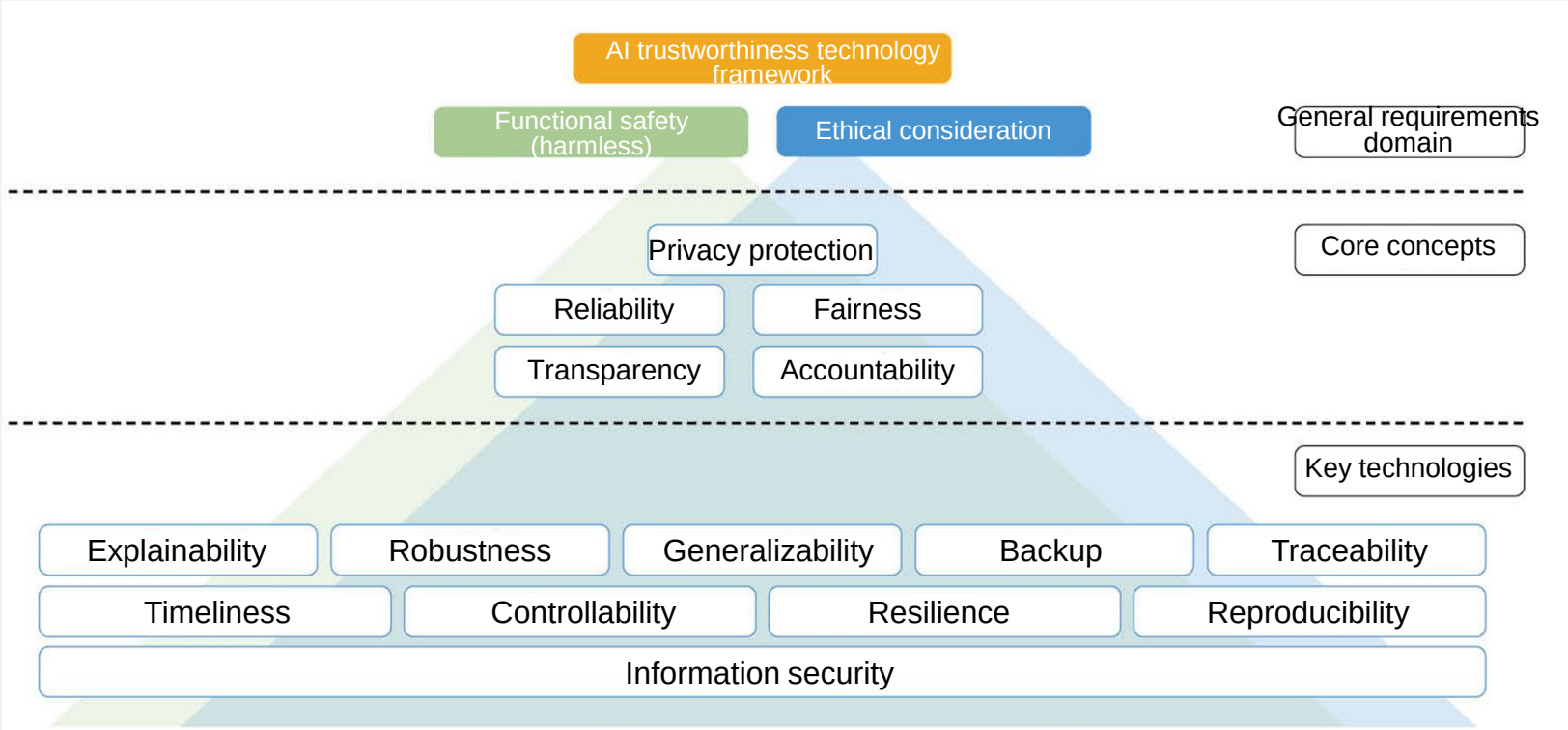


2023 Edition

Standardization frameworks for AI Trustworthiness

An White Paper on Standardization for AI Trustworthiness

- Describes the state of the art with regard to policies, laws and regulations, and standards related to AI trustworthiness inside and outside China
- Proposes technology and standardization frameworks for AI trustworthiness
- Categorizes trustworthiness related technologies into general requirements, core concepts, and key technologies
- Systematically describes the problems that AI trustworthiness needs to address



AI Trustworthiness technology framework

- **General requirements**

Proposes basic standards for AI trustworthiness in aspects such as technology architecture, functional safety, and ethical consideration; sets out the technical architecture, concepts, and dependencies; puts forward framework requirements for functional safety.

- **Core concepts**

Introduces the core concepts pertaining to AI systems. These concepts are relatively independent and targeted. They are suitable for establishing overarching guidelines and providing benchmarks and yardsticks for products, systems, and services in the AI industries.

- **Key technologies**

Puts forward the key technology requirements for AI systems. Each element is relatively focused. The proposition includes independent and systematic test method theories and indicators to facilitate the development of evaluation standards for specific practices.

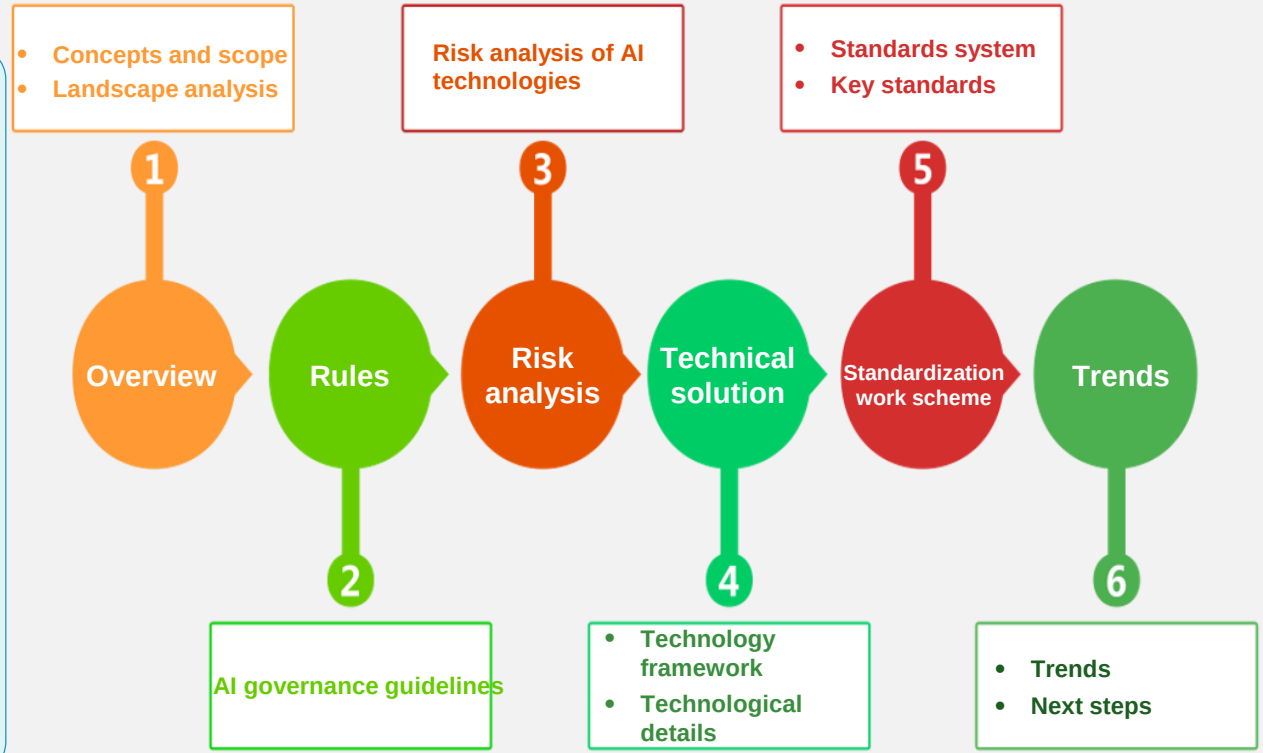
Guidelines for the standardization of AI governance(2023 Q1)

Background: Several sets of policies have been released in China, including *Opinions on Strengthening the Ethical Governance of* Science and Technology, *Ethics Code for New-Generation of AI* . Currently, research is focused on four areas related to AI governance: **concepts and scope**, **risk assessments**, **governance technologies**, and **governance standardization**.

Research focus

To support the development of standards, the research:

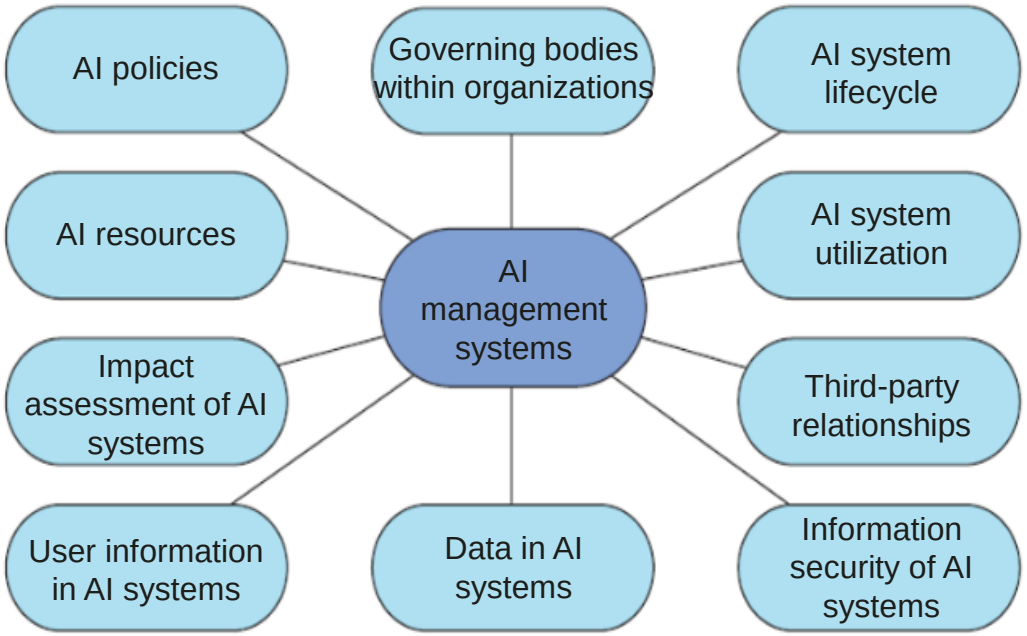
- **Clarifies the concepts and scope of AI governance**
- **Refines both the essential and extended meaning of the AI governance principles**
- **Evaluates and classifies risks**
- **Puts forward a technology framework for AI governance**
- **Further explores the standardization system for AI governance**
- **Proposes next steps for AI governance**



Prioritized Tasks for Standardization of AI Governance

Priorities for 2023:

- Key task 1: Continue to research AI governance, promote projects such as the ***White Paper on Standardization of Artificial Intelligence Algorithm Governance***, and research key elements such as fairness, explainability, and privacy.
- Key task 2: **Accelerate the development of a standardized AI governance system**, and promote the establishment of national standards such as the AI **trustworthiness standard**, the **AI management system**, and the **AI Risk Management** standards.



- Key task 3: Improve the **basic service platform for AI assessment and testing**, select relevant high-quality case studies, and promote the healthy development of the industry.

Examples of the AI trustworthiness related standards in progress

Type	Title	Scope	Orgs	Status
National standards	**Artificial intelligence - Trustworthiness - Part 1: General**	Technical framework, requirements and evaluation processes of AI trustworthiness	SAC/TC 28/SC 42	NP
	Artificial Intelligence - Risk Management Capability Assessment	Risk elements, risk management requirements, and capability assessment and processes		
	Artificial Intelligence - Management System	Requirements and guidance for establishing, implementing, maintaining and continually improving an AI management system		AWI
National standards	**Information security technology - Generative AI manual annotation security specification**	Processes, guidance, security control and testing aspects of manual annotation for generative AI	SAC/TC 260	NP
	Information security technology - Generative AI pretraining and refinement training data security specification	Training data classification framework, content risk level, processing security requirements and verification methods		
	Information security technology - General security requirements for generative AI	The overall requirements that generative AI services need to follow in terms of security		
	Information security technology - AI computing platform security framework	Security functions, mechanism, modules and services API of AI computing platform security framework		CD
Industry standards	**AI medical devices - Quality requirements and evaluation**	Series of vocabulary, dataset, data labelling and trackability aspects of AI medical devices	NIFDC	Published

Huawei AI Business Intent and Governance Principles

AI Governance Principles

For People

1. Human-AI Collaboration，Serve People
2. Transparency

For Society

3. Diversity, Non-discrimination & Fairness

For Product Quality

4. Safety & Robustness
5. Data Security & Privacy
6. Traceability & Accountability

Basic Thought1:

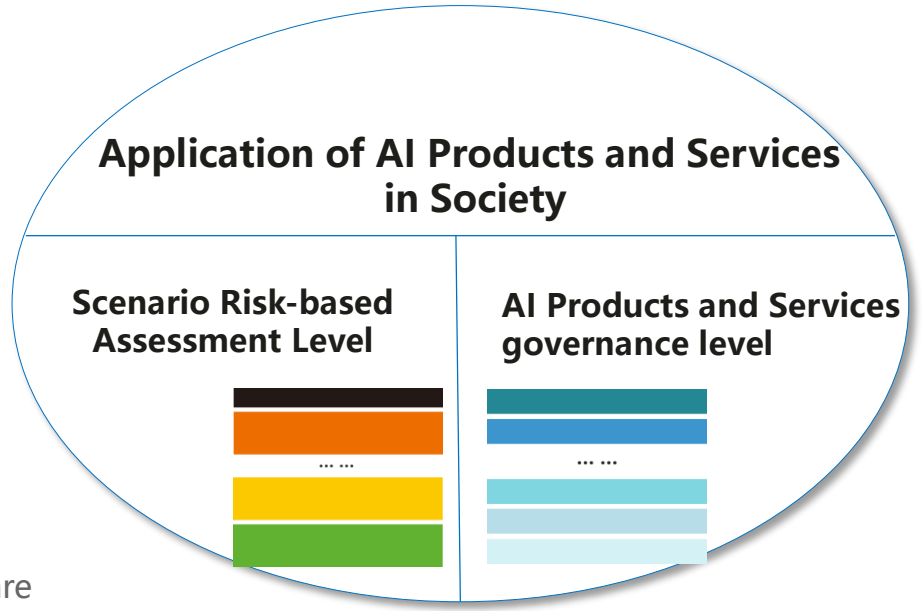
Establish a governance mechanism with joint efforts from five roles

◆ Regulatory Authority

Empowered by the public to supervise the other parties in the mechanism to achieve governance requirements

◆ Public

Applied or affected persons in AI application scenarios or to which persons who are not directly related, but whose interests may be harmed in the future



◆ User

Users (e.g. operator/deployer/purchaser) are organizations or individuals that bring AI products and services into life and production practices for specific purposes.

◆ Product or Service Provider

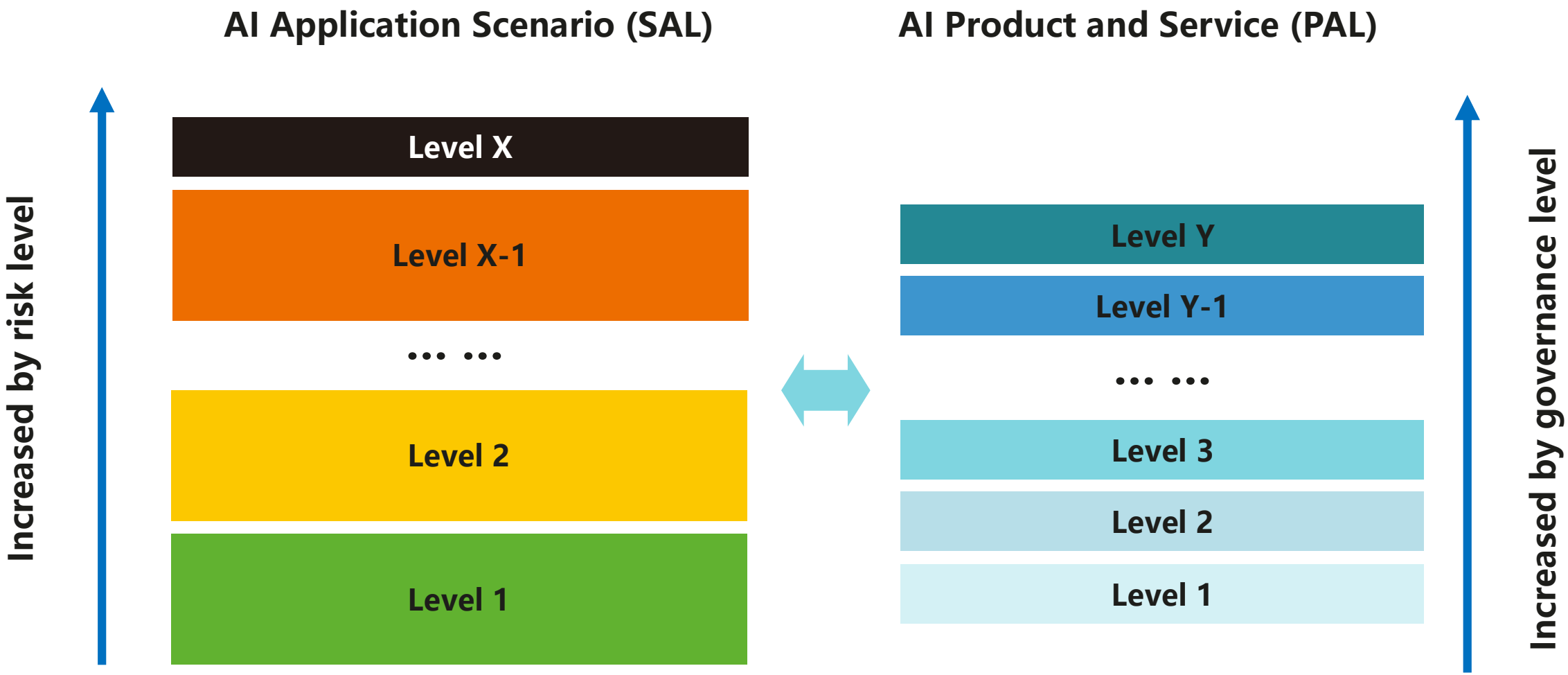
Provide AI products or services for the society, which satisfy the requirements from various stakeholders.

◆ Certification/Audit Body

Neutral Part, balance for ecosystem, provide certification or audit services for providers or users, recognized and authorized by Regulatory Authority, set up confidence for Public

Basic Thought2:

There should be a mechanism to coordinate five-roles, towards a neutral 3rd-party certification

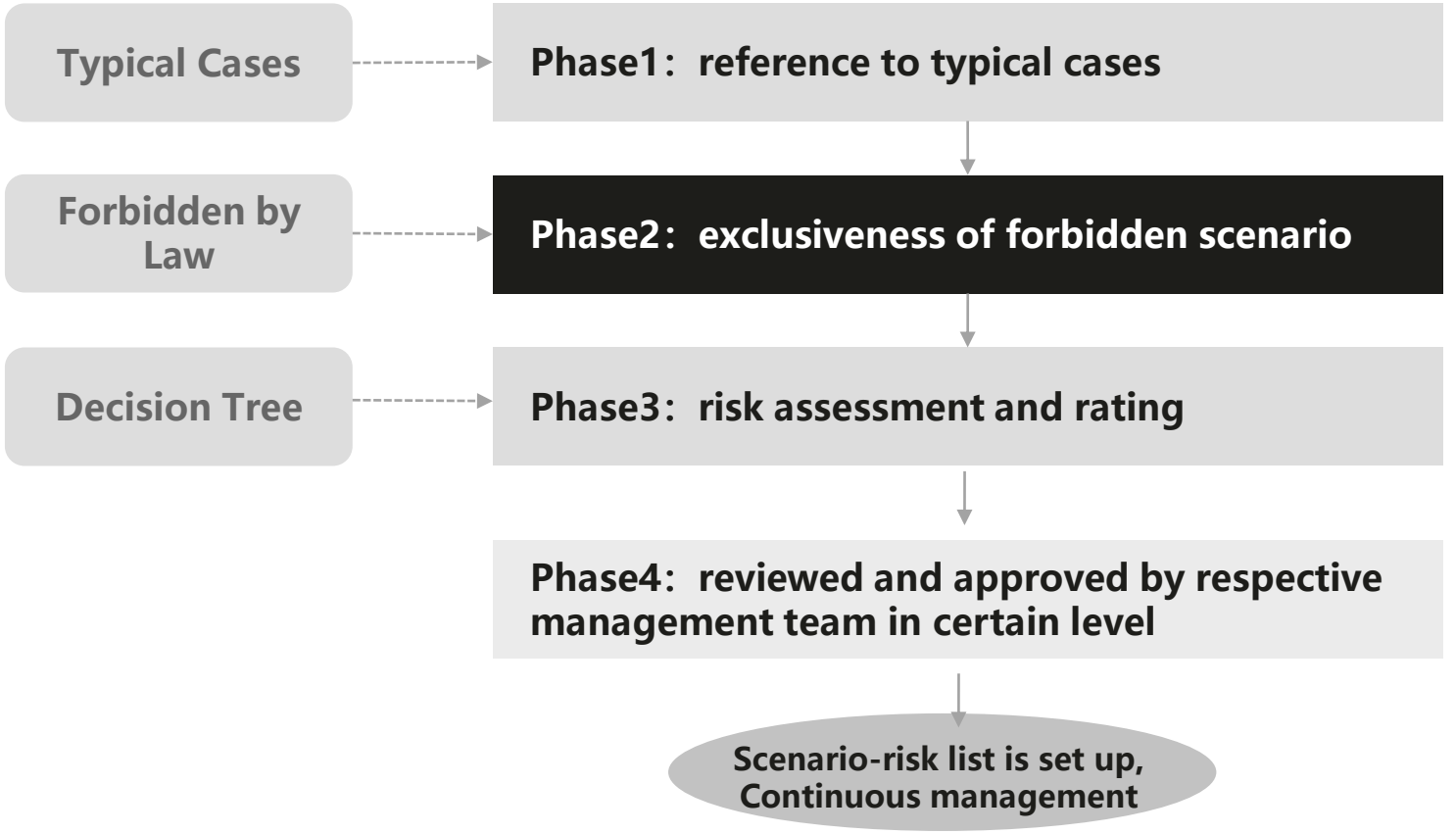


Trial of risk assessment and rating based on scenario, to identify and manage risk effectively

Risk assessment and rating method guideline

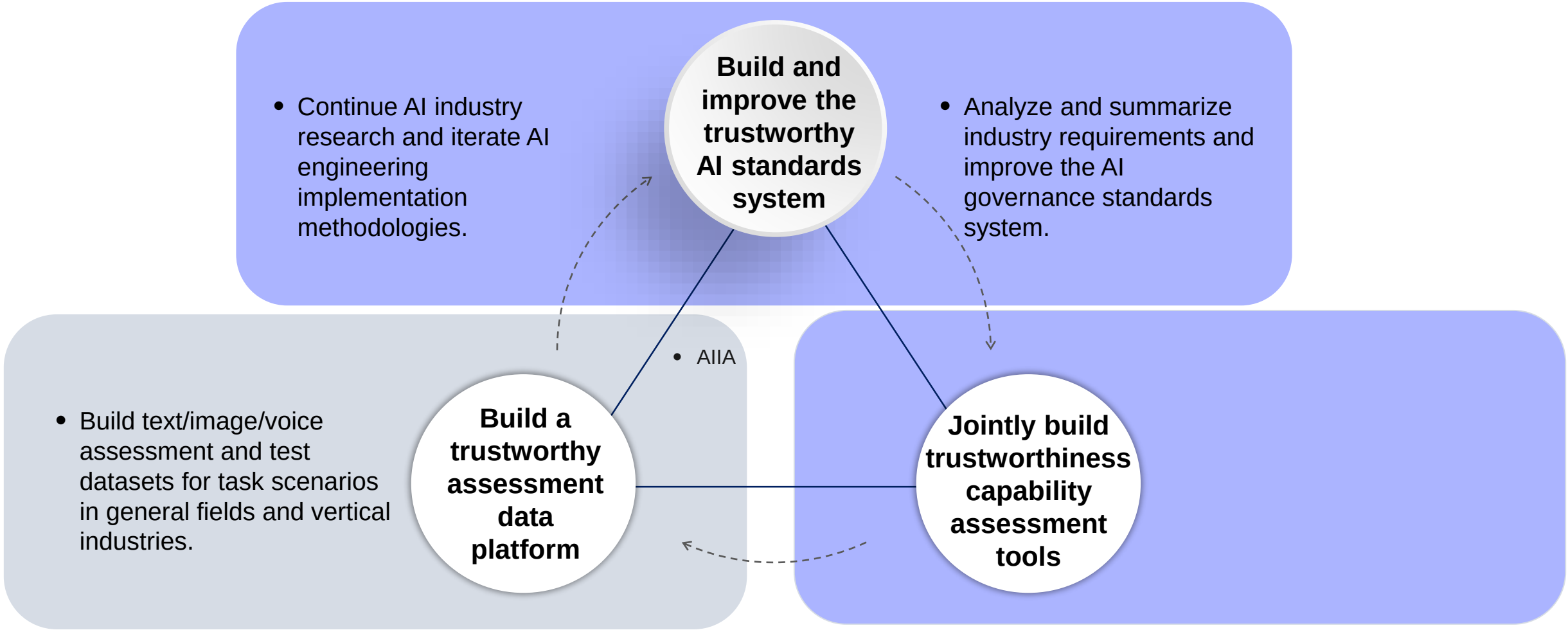
(For illustration)

Process for risk assessment and rating based on scenario



As a Industrial Alliance, AIIA takes joint effort to explore the effective assessment method of AI Governance

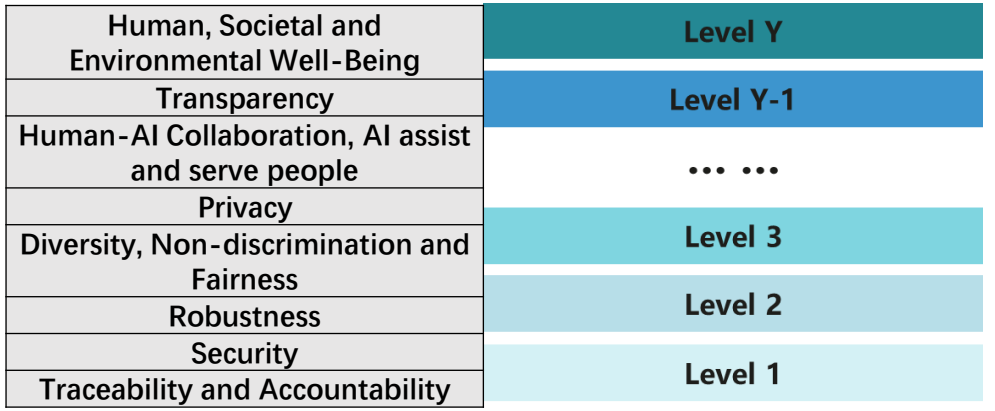
Artificial Intelligence Industry Alliance (AIIA) continue to facilitate trustworthy AI standardization by **engineering means** to ensure the sustainable development and trustworthiness of AI technologies and applications.



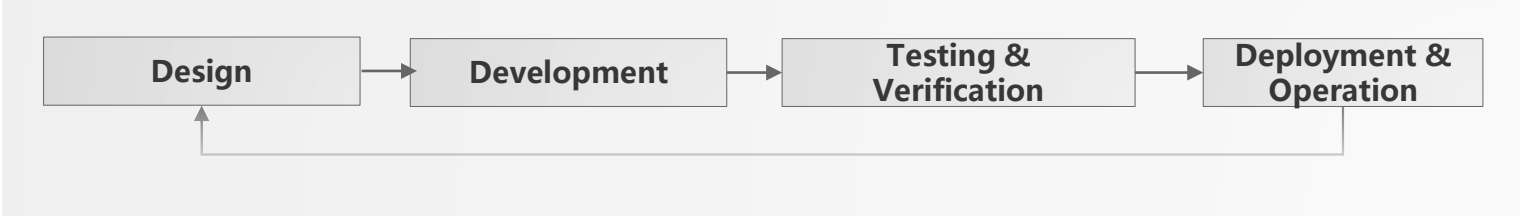
We are still on the road, a Agile Governance is needed

Challenge1: How to assess and rate risk objectively, recognized by all roles

Challenge2: How to transform abstractive principles & objectives to measurable criteria & indicators



Challenge3: How to achieve the governance goals by certain engineering practices



Thank You