Translation

The following white paper from a Chinese think tank describes the state of artificial intelligence in China and the world. It divides its focus among AI innovation and breakthroughs, engineering and other practical uses of AI, and AI governance initiatives in the areas of trustworthiness and safety.

Title
Artificial Intelligence White Paper (2022)
人工智能白皮书（2022年）

Author
The China Academy of Information and Communications Technology (CAICT; 中国信息通信研究院; 中国信通院). CAICT is a think tank under the PRC Ministry of Industry and Information Technology (MIIT; 工业和信息化部; 工信部).

Source
CAICT website, April 12, 2022.

The Chinese source text is available online at:
http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf
An archived version of the Chinese source text is available online at: https://perma.cc/SAC8-GZ5D
U.S. $1 ≈ 6.7 Chinese Yuan Renminbi (RMB), as of June 6, 2022.

Translation Date	**Translator**	**Editor**
June 6, 2022	Etcetera Language Group, Inc.	Ben Murphy, CSET Translation Manager

CAICT 中国信通院

No. 202205

Artificial Intelligence White Paper

(2022)

China Academy of Information and Communications Technology
(CAICT)

April 2022

Copyright Statement

Preface

Against the background of the new technological and scientific revolution and industrial transformation, the deep integration of artificial intelligence (AI) and industry is an inevitable choice to release the multiplication effect of digitalization, accelerate the development of strategic emerging industries, and build comprehensive competitive advantages. At present, the rapid penetration of AI into various industries is promoting cross-cutting integration and development among emerging industries, among emerging industries and traditional industries, and between technology and society. At the outset of the "14th Five-Year Plan," comprehensively sorting out the development trends of AI is of great relevance and significance.

This white paper focuses on providing a description from the dimensions of AI policy, technology, applications, and governance. At the **policy level**, the strategic position of AI has been continuously strengthened in China and abroad to promote the release of AI dividends. At the **technical and application level**, AI technology represented by deep learning has developed rapidly, and new technologies have begun to be explored and applied; engineering capabilities (工程化能力) have been continuously enhanced and continue to be deeply applied in fields such as healthcare, manufacturing, and autonomous driving; and trustworthy AI technology has attracted widespread attention from society. At the same time, **governance-level** work has also received a high level of attention from the world. The regulatory process continues to accelerate in various countries, and industrial practices based on trustworthy AI continue to deepen.

On the whole, this White Paper holds that AI has gradually entered a new stage, and the development direction in the next step will be defined and guided by the "three-dimensional" (3D) coordinates of **technological innovation, engineering practices** (工程实践)**, and trustworthiness and safety** (可信安全). Specifically, the first dimension highlights innovation. Innovations centered around algorithms and computing power (compute) will continue to constantly emerge. The second dimension highlights engineering. Engineering capabilities have gradually become the key factor that allows AI to empower thousands of industries on a large scale. The third dimension highlights trustworthiness. The development of responsible and trustworthy AI has become a consensus, and the implementation of abstract governance principles throughout the AI lifecycle process will become the focus.

Due to the rapid development of AI, its wide range of influence, and its unprecedented degree of influence, we need to further deepen our understanding of AI. We welcome your criticisms and corrections for any shortcomings in this White Paper.

Contents

List of Figures

I. Overview of the Development of AI

AI is an emerging strategic technology that will guide the future and an important driving force for the new round of science and technology (S&T) revolution and industrial transformation. Many times, General Secretary Xi Jinping has given important instructions, emphasizing that "we must deeply grasp the characteristics of the development of new generation AI, strengthen the integration of AI and industry development, and provide new momentum for high-quality development." In recent years, AI-related technologies have continued to evolve, the process of industrialization and commercialization has continued to accelerate, and their in-depth integration with thousands of industries is currently accelerating. Standing at the special stage of the beginning of the "14th Five-Year Plan," we firmly believe that a comprehensive review of the development trend of AI policies, technologies, applications, and governance can help build consensus in the industry and promote the sustainable and healthy development of AI.

(i) All Countries Are Constantly Upgrading AI Strategies and Seizing Important Development Opportunities One After Another

AI has become a key field of technological innovation and an important pillar of the digital economy era. Since 2016, more than 40 countries and regions have raised the development of AI to the level of national strategy. In the past two years, especially under the impact of the COVID-19 pandemic, more and more countries have recognized that AI plays a key role in enhancing global competitiveness and have deepened their AI strategies one after another. The **EU** released the *2030 Digital Compass: the European way for the Digital Decade* and *Updating the 2020 Industrial Strategy*, which intend to comprehensively reshape the global influence of the digital age. In these documents, the promotion of AI development is listed as an important task. The **United States** successively established the National Artificial Intelligence Initiative Office, the National Artificial Intelligence Research Resource Task Force, and other institutions. Various departments have intensively issued a series of policies, raising AI to the height of an "industry of the future" and "technology of the future," constantly consolidating and enhancing the global competitiveness of the United States in the field of AI, and ensuring its "bellwether" status. Following the formulation of its *Integrated Innovation Strategy 2020*, **Japan** released the *AI Strategy 2021* in June 2021, which is committed to promoting innovation and creation plans in the field of AI and comprehensively building digitized government. In September 2021, the **United Kingdom** released a new ten-year strategy for national AI, which is another important strategy launched after 2016 and aims to reshape the influence of the AI field. **China's** *Proposal of the Central Committee of the Chinese Communist Party on*

Drawing Up the 14th Five-Year Plan for National Economic and Social Development and Long-Range Objectives for 2035[1] points out that we must target cutting-edge fields such as AI, implement a number of forward-looking and strategic major S&T projects, and promote the healthy development of the digital economy.

Investments to meet innovation needs in the AI field continue to increase. The promotion of AI development through incentive programs and direct investment projects is already a widespread practice of major economies. The EU continues to increase financial support for the AI industry and vigorously promote digital transformation in Europe. The EU's largest-ever project to support R&D and innovation, the "Horizon Europe" program, has a total investment of 95.5 billion euros and explicitly includes AI in the scope of financial support. In April 2021, in the form of regulations, the EU used the "Digital Europe Programme" to invest in projects including AI, with a total of 7.59 billion euros. **The United States sees maintaining its leading position as its strategic goal and continues to increase investment in the AI field.** The U.S. non-defense budget for AI in 2021 increased by about 30%, to a total of U.S. $1.5 billion. In addition, in the *United States Innovation and Competition Act*, AI, quantum computing, and other fields are listed as priorities in the U.S. R&D budget for fiscal year 2022. In the future, a total of U.S. $100 billion will be invested in R&D in various fields including AI. **The UK regards investing in and planning the AI ecosystem as a long-term strategy**, launching a national AI research and innovation program and supporting advanced AI research. According to statistics, between 2014 and 2021, its investment in AI exceeded 2.3 billion pounds.

Using applications to lead and promote the implementation of AI technology has become the consensus of various countries. The United States guides the innovation and integrated application of AI technology in industries and sectors. In July 2021, the U.S. National Science Foundation partnered with various departments and well-known enterprises to establish 11 new national AI research institutes, covering human-computer interaction, AI optimization, dynamic systems, reinforcement learning, and other research directions, with research projects covering multiple fields, such as construction, healthcare, biology, geology, electricity, education, and energy. **The UK supports AI industrialization,** launching a joint plan between the Office for Artificial Intelligence and UK Research and Innovation to ensure that AI benefits all

[1] Translator's note: For an English translation of the CCP Central Committee Proposal on the 14th Five-Year Plan, see: https://cset.georgetown.edu/publication/proposal-of-the-central-committee-of-the-chinese-communist-party-on-drawing-up-the-14th-five-year-plan-for-national-economic-and-social-development-and-long-range-objectives-for-2030/. For an English translation of the final, authoritative version of China's 14th Five-Year Plan, see: https://cset.georgetown.edu/publication/china-14th-five-year-plan/.

industries and regions and promote the widespread application of AI. **Japan focuses on infrastructure construction and AI applications,** proposing to accelerate the construction of relevant infrastructure, emphasizing cross-industry data transmission platforms and AI-related standards, comprehensively promoting the application of AI in various industries such as healthcare, agriculture, transportation and logistics, smart cities, and manufacturing, and increasing support for small and medium-size enterprises (SMEs). The outline of China's 14th Five-Year Plan[2] clearly states that vigorously developing the AI industry, building AI industry clusters, and the in-depth empowerment of traditional industries will be focus points. In April 2021, the Ministry of Industry and Information Technology supported the creation of the second batch of national AI innovation and application pilot zones in Beijing, Tianjin (Binhai New Area), Hangzhou, Guangzhou, and Chengdu and continuously strengthen the guiding role of applications. The Ministry of Science and Technology will support the construction of a number of AI innovation and development pilot zones and successively approve 15 national new generation AI innovation and development pilot zones in Beijing, Shanghai, Tianjin, Shenzhen, Hangzhou, and other regions.

(ii) AI Has Entered a New Stage with Sustainable and Healthy Development Becoming the Focus

Since the birth of AI in 1956, related theories and technologies have continued to evolve. It was only in the past decade that AI has been able to truly move from laboratory research to industrial practice on a large scale. This was due to breakthroughs in deep learning and other algorithms, the continuous increase in computing power, and the continuous accumulation of massive data (海量数据). In the process of industrial development and empowerment, on the one hand, a large number of practical scenarios can see the development path from "able to use" to "easy to use." This is inseparable from the continuous iteration of the technology itself, the continuous optimization of engineering implementations, and the support and guarantees provided by management systems. On the other hand, with the exposure of various risks and challenges in AI applications and the continuous deepening of people's understanding of AI, AI governance has become a topic receiving a high degree of attention from all sectors around the world, and voices calling for trustworthiness and safety are continuing to increase.

In addition to attaching importance to AI technological innovation in the future, we must also pay more attention to engineering practices and trustworthiness and safety. This also constitutes new "3D" development

[2] Translator's note: CSET's English translation of China's 14th Five-Year Plan Outline is available online at: https://cset.georgetown.edu/publication/china-14th-five-year-plan/.

coordinates, leading the AI technology industry to a new stage. In fact, the industry's efforts in various dimensions have already begun and have never stopped, but today, engineering practices and trustworthiness and safety have been placed in a more important position. The 3D coordinates are not completely independent, but interwoven and mutually supporting. Figure 1 shows a schematic of the evolution of the current wave of AI along various directions and outlines the development context under each coordinate.

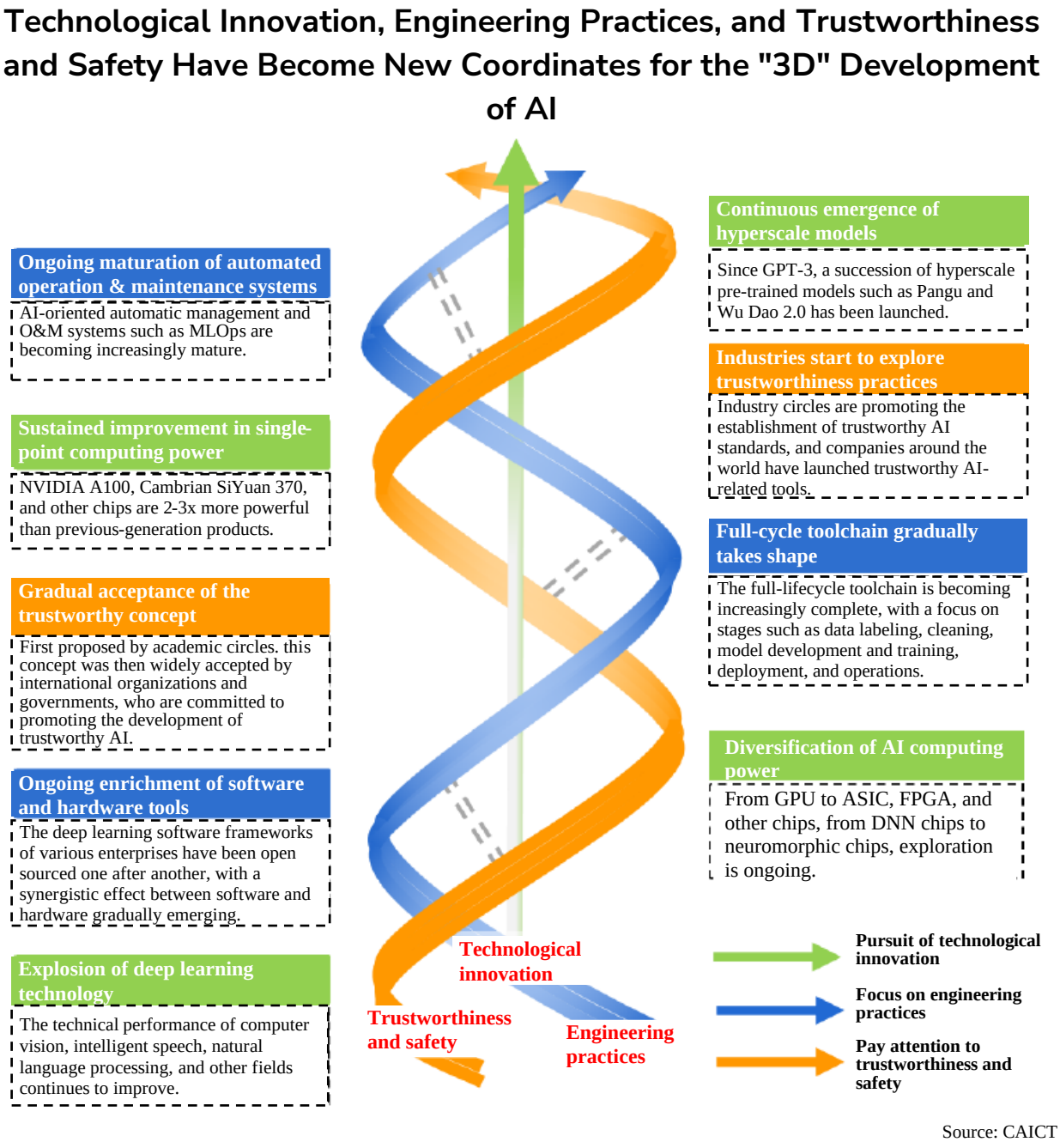


Figure 1 Schematic of the Three Dimensions of AI Evolution

The pursuit of technological innovation in specific scenarios has always been the goal and driving force of AI development. The explosion of algorithms as represented by deep learning opened the curtain on the AI tidal wave. AI is now widely used in computer vision, intelligent speech, natural language processing, and other fields, where it has successively surpassed the level of human recognition. The diversification of AI computing power and the continuous improvement in single-point computing power (单点算力) have provided strong support for the development of AI.

Recently, hyperscale pre-trained models have emerged frequently in China and abroad, constantly refreshing the ranking list of application fields. In the future, changes in algorithms and computing power will continue, laying the foundation for an era of greater intelligence.

Engineering practices and capabilities have increasingly become an important support for releasing the dividends of AI technology. Efforts in engineering practices can be traced back to the birth of open-source frameworks such as Caffe, TensorFlow, and PaddlePaddle. By shielding the underlying hardware and operating system details, these frameworks greatly reduce the difficulty of model development and deployment, effectively promoting the proliferation of AI technology. At present, the integration of AI with supporting technologies such as cloud computing and big data continues to deepen, and the toolchains centered on various links such as data processing, model training, deployment and operations, and security monitoring are constantly enriched. The AI R&D management system is becoming increasingly complete, and automated operations and maintenance (O&M) technology, as represented by MLOps, has received more and more attention. With the continuous improvement of engineering practices and capabilities, the empowerment method of the "small workshops and project system" (“小作坊、项目制”) is becoming a thing of the past. In the future, it will be more convenient and efficient to realize AI applications and product delivery.

Trustworthiness and safety have gradually become an indispensable guarantee in the process of AI empowerment. Trustworthy AI was first proposed by academic circles. In recent years, research on trustworthy AI centered on safety, stability, explainability, privacy protection, and fairness has continued to heat up. The concept of trustworthy AI has received widespread attention from international organizations. The "G20 AI Principles," proposed by the Group of Twenty (G20) in June 2019, clearly propose to promote the development of innovations in trustworthy AI, which has become an important consensus. The concept of trustworthy AI has been gradually implemented throughout the full life cycle of AI and industrial practices have been continuously enriched. This has evolved into an important methodology for implementing the relevant requirements of AI governance.

In general, AI is entering a **new stage of "being innovation-driven, deepening applications, development of norms."** From the perspective of the industrialization of AI itself, iterative technological upgrades are the source of development. At present, AI is not perfect, the path of intelligentization (智能化) is still being explored, and the innovation-drivenness of technology will help to open up new space for development. From the perspective of the empowerment of traditional industries by AI, especially since the pandemic, the digitalization and intelligentization transformations have been

accelerating, pushing AI applications onto the fast track, while related applications have continued to deepen. From the governance perspective, technology and industrial development necessarily run ahead of regulations and systems. Governance issues are becoming more and more serious, and ensuring the healthy development of AI has become a global concern. Here, there are both gradual changes and structural and even directional adjustments. It is necessary to improve capabilities comprehensively and systematically in all respects so as to promote the sustainable and healthy development of AI.

II. AI Technology and Applications Continue to Evolve Along the Three Directions of "Innovation, Engineering, and Trustworthiness"

In the new context, AI technology also needs to adapt to new changes. This chapter untangles the development trends of AI technology and applications according to the new 3D coordinates. Technological innovation centered around algorithms, compute, and data is always the dominant theme of progress. Relevant technologies in engineering practices have started to cover the whole process of AI, accelerating large-scale AI applications. Trustworthy AI technology is an important support for solving governance difficulties and has attracted increasing attention from all sectors.

(i) AI Continues to Make Breakthroughs in the Pursuit of Extreme Innovation

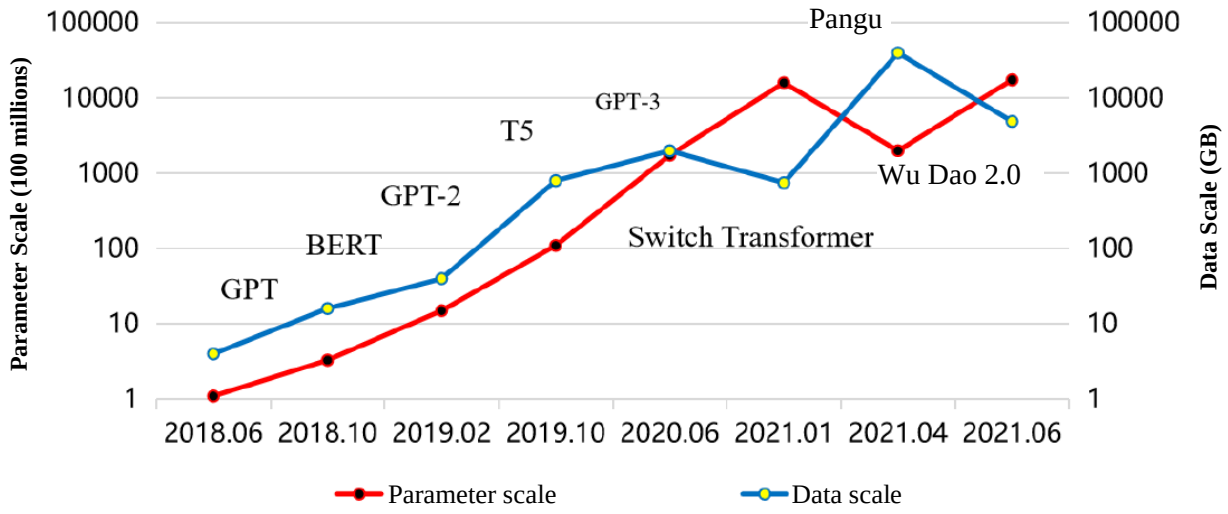
Algorithms, compute, and data have always been regarded as the three horses that pull AI development and an important basis for promoting AI development. **At the algorithm level,** hyperscale pre-trained models have become one of the hottest topics in the past two years, and the ranking lists in various fields are constantly being refreshed. Research on knowledge-driven AI and other directions has become an important means to explore the improvement of cognitive abilities. The integration and innovation of AI and various scientific research fields has attracted increasing attention, and AI has become an important tool in basic scientific research. **At the level of basic computing power,** single-point compute continues to increase, and the customization and diversification of computing power has become an important development trend. Computing technology centered around the three major capabilities of data processing, data storage, and data interaction has evolved and been upgraded, with continued exploration in areas such as neuromorphic chips and quantum computing. **At the data level,** AI technology, represented by deep learning, requires a large amount of labeled data. This has also spawned specialized technologies and even services. With the continuous increase in the specificity and depth of problem-orientation, data services have become refined and customized. In addition, as the importance of knowledge in AI is widely discussed, the construction and utilization of knowledge sets (知识集)

continues to increase.

1. New Algorithms Constantly Emerge, and Technology Integration Constitutes an Important Trend

Hyperscale pre-trained models promote the continuous improvement of technical performance and continue to develop in the direction of larger scales and more modalities. Since OpenAI launched GPT-3 in 2020, companies and research institutions such as Google, Huawei, Beijing Academy of Artificial Intelligence (BAAI), the Chinese Academy of Sciences (CAS), and Alibaba have successively launched hyperscale pre-trained models, including Switch Transformer, DALL-E, MT-NLG, Pangu, Wu Dao 2.0, Zidong Taichu (紫东太初), and M6, constantly refreshing various ranking lists. The General Language Understanding Evaluation (GLUE) comprehensive score of Baidu's ERNIE 3.0 model exceeded 90% on natural language understanding tasks. Compared with OpenAI's CLIP model, the multi-source image and text dataset score (RUC-CAS-wenlan) of BAAI's Wu Dao – Wen Lan model showed a significant improvement of 37.0%. At present, the number of pre-trained model parameters and the scale of training data are increasing at a rate of 300x/year, continuing through the short-term evolution direction of increasing the size of models and increasing the scale of training data. In addition, the cross-modal pre-training of large models is becoming increasingly common. This has transitioned from learning only text data in the early days to jointly learning text and images, with it now being possible to process tri-modal data with text, images, and speech. In the future, pre-trained models that use more image encodings, more languages, and more types of data will emerge. This will also be beneficial to the exploration of AI generalization.

Rapid Growth in Parameter Quantity and Training Data Scale of Large Models



Source: CAICT

Figure 2 Schematic of the Growth of Large Model Parameters and Training Data Scale

Lightweight deep learning technology continues to be explored, and computing efficiency has significantly increased. Complex deep learning models often consume a large volume of storage space and compute and are difficult to apply in resource-constrained conditions such as on terminals and in edge computing. Technologies with the advantages of low memory and low computational effort have become industry necessities. Lightweight deep learning has become an important technology for solving this challenge. This includes research directions such as designing more compact and efficient neural network structures, pruning large models (that is, "cropping" out part of the model structure), and quantizing network parameters to reduce the calculation volume. For example, MobileNet proposed by Google and ShuffleNet proposed by Megvii have become typical representatives of compact models. The lightweight PaddleOCR model launched by Baidu was reduced in size to 2.8MB and became popular after being open sourced on GitHub.

"Generative AI" technology continues to mature, and in the future, the abilities of listening, speaking, reading, and writing will be organically combined. At present, "generative AI" technology is widely used in intelligent writing, code generation, audio reading, news broadcasts, voice navigation, image restoration, and other fields. The automated synthesis of text, speech, images, videos, and other data by machines is driving a revolution in the production of digital content on the Internet. The organic combination of listening, speaking, reading, writing, and other abilities has become the development trend of the future. For example, CCTV, Xinhua News Agency, and Gmw.cn have launched digital human anchors, which support the one-click generation of video from audio/text content and can achieve the rapid and automatic production of program content. Relevant digital human anchors and digital human reporters have been widely used for large-scale reports and programs, such as the two sessions[3] and the Spring Festival Gala.

Knowledge computing has become an important field of exploration for promoting the transformation of AI from perceptual intelligence to cognitive intelligence. Knowledge condenses human wisdom, and the dual drivers of knowledge and data help to solve the problem of inference and decision-making with incomplete information, uncertainty, and a dynamic environment. This can improve the level of intelligence of AI systems. At present, centered around knowledge acquisition, knowledge modeling, knowledge management, knowledge application, and other

[3] Translator's note: The "two sessions" (两会) are the annual full sessions of the National People's Congress (NPC; 全国人民代表大会; 全国人大), China's parliament, and the National Committee of the Chinese People's Political Consultative Conference (CPPCC; 中国人民政治协商会议全国委员会; 全国政协), an advisory body, which are held concurrently each year in March.

processes, technologies covering knowledge graphs, knowledge bases, and graph computing have already been formed. This system covering knowledge representation, knowledge computing, knowledge inference, and decision-making capabilities can achieve knowledge management and utilization. Both academic and industry circles have begun to launch knowledge-based AI application platforms or solutions. For example, Tsinghua University, Zhejiang University, Huawei Cloud, BAAI, Baidu, Emotibot, and Gridsum have launched solutions such as knowledge computing engines, knowledge middle grounds (知识中台), knowledge engineering platforms, and knowledge intelligence platforms. Going forward, knowledge computing will focus on embedding prior knowledge in deep learning algorithms to build explainable models. This way, knowledge can be deeply involved in model solving, further improving the efficiency and quality of, and the robustness, explainability, and transferability of, AI.

The integration of AI and scientific research continues to deepen, which has begun to "subvert" the traditional research paradigm. In recent years, the ability of AI to analyze vast quantities of data has freed researchers from being limited to conventional "deriving theorem"-style (“推导定理式”) research. Instead, they can find relevant information based on high-dimensional data to accelerate the research process. In 2020, DeepMind proposed AlphaFold2, which won the first prize in the Critical Assessment of protein Structure Prediction (CASP). It could accurately predict the 3D structure of proteins, with an accuracy comparable to the 3D structure solved using experimental techniques such as electron cryomicroscopy. The Sino-U.S. research team used AI methods to increase the limit on molecular dynamic [simulations] by several orders of magnitude while ensuring "ab initio" high precision. Compared with similar work in the past, the computing space scale is increased by 100x, and the calculation speed was accelerated by 1000x. The team won the 2020 ACM Gordon Bell Prize. Even more surprising, integrated explorations of AI with the mechanics, chemistry, materials science, biology, and even engineering fields continue to emerge, and the depth and breadth of AI applications will continue to expand in the future.

2. Breakthroughs Continue in Single-Point Computing Power, and New Technologies Are Still in the Exploratory Stage

At present, breakthroughs continue to be made in AI computing power, and chips for training and inference are still evolving rapidly. This is mainly driven by the demand for compute. On the one hand, this is reflected in the model training stage. According to OpenAI data, the growth rate of model computing power far exceeds the growth rate of AI hardware computing power, with a 10,000-fold gap between the two. On the other hand, due to the ubiquity of inference, the demand for computing power for inference continues to grow. At the same time, research continues on new

computing power architectures. Neuromorphic chips, in-memory computing, and quantum computing have attracted much attention, but they are generally in the exploratory stage.

Innovation in training chips has accelerated, and inference chips are developing in the direction of dedicated customization. GPU-based training chips have continued to increase, and companies oriented toward GPU innovation have begun to exert their capabilities. A group of startups focusing on the GPU track has emerged, including Moore Threads, Iluvatar CoreX, and Biren Technology. **The capabilities of cloud training chips based on ASIC and other architectures have significantly improved.** Cambricon's SiYuan 370, Enflame Technology's "DTU 2.0 (邃思 2.0)," and Baidu's Kunlun II have increased computing power by a factor of 3-4 compared to the previous generation. **Dedicated custom end-to-end inference chips are emerging everywhere, and smart chips for mobile phone applications have become a highlight.** In January 2021, MediaTek launched the high-end mobile phone chip Dimensity 1200, which can process 5G, AI, and image data at the edge. In August, Google launched its first smartphone chip, Tensor, exclusively for its Pixel line of phones.

Neuromorphic chips, in-memory computing, and quantum computing are still the key directions of exploration. Neuromorphic chips, in-memory computing, quantum computing, and other technologies can achieve the advantages of high compute and low power consumption at the theoretical level. Although some progress has been made, in general, their current technological maturity is relatively low. The Center for Brain Inspired Chips at Peking University announced achievements such as the "Ultra-low-power Smart IoT Chip (AIoT)" at ISSCC in 2021. New AI chips are favored by investment funds. Since 2021, many companies have completed A round or A+ round financing for hundreds of millions of Chinese yuan Renminbi (RMB), including 3D vision AI chip manufacturer Aivatech, Reexen, which focuses on the research and development of neuromorphic sensor storage and computing integrated chips, and the AI vision chip R&D company Axera (爱芯科技).

3. The Continued Increase in Data Scale Builds a Hotspot for Domain Knowledge Integration

The rapid development of AI promotes the continuous increase in the scale of data. According to IDC estimates, the global data scale will reach 163ZB in 2025, with unstructured data accounting for 80%-90%. Data services have entered the stage of deep customization. Baidu, Alibaba, JD, and other companies have launched data customization services based on different scenarios and needs. The data sets required by enterprises are transitioning from general purpose simple scenarios to personalized complex scenarios, such as the progression of speech recognition data sets from

Mandarin to less-widely spoken languages, dialects, and other scenarios and the progression of intelligent dialogue datasets from scenarios such as short-answer Q&A and voice control to application scenarios and business Q&A.

All parties are actively exploring the establishment of high-quality knowledge sets to support the future development of knowledge-driven AI applications. A knowledge set contains traditional data such as voice, image, and text as well as definitions, rules, logical relationships, and other data. It is a data-based presentation of knowledge. Well-known knowledge sets in industry include Wordnet and Hownet. For example, Alibaba and Hong Kong Polytechnic University developed the FashionAI knowledge set based on clothing design knowledge, which accelerated the application of AI in the clothing design industry.

(ii) The AI Toolchain Has Become the Core of Engineering Practices and Capabilities

With the continuous development of AI technology, engineering applications have accelerated in recent years. In the financial field, AI technology has begun to deeply penetrate the whole process of front, middle, and back offices. Medical AI has begun to enter the marketization stage. As of the end of August 2021, a total of 28 products had been approved for Class III medical device registration certificates. With the rapid development of AI in the manufacturing field, Deloitte predicts that China will maintain an average annual growth rate in excess of 40% over the next five years. At present, the application of AI by enterprises is showing a transition from preliminary exploration to large-scale application. Generally speaking, the continuous improvement of engineering practices and capabilities has become the key to future applications.

AI engineering has started to become a focus of attention from all sectors. In academic circles, the Software Engineering Institute of Carnegie Mellon University has launched AI engineering research in recent years and has undertaken a national research program funded by official U.S. institutions in conjunction with universities and industry. World-renowned AI experts Michael I. Jordan and Eric Xing (刑波) believe that AI engineering is an emerging engineering discipline and a trend in the development of AI from a theoretical discipline to an engineering discipline. In industry circles, Gartner has listed AI engineering as one of its annual strategic technology trends for two consecutive years. Alibaba Cloud and other enterprises regard AI engineering as the key to transforming AI into enterprise productivity.

AI engineering focuses on the efficient coupling of the full life cycle process of tool systems, development processes, and model management. At the tool system level, systemization and openness have become the development

characteristics of the R&D platform technology toolchain. A relatively complete tool system has been initially built centered around technologies such as machine learning and deep learning. This system greatly reduces the difficulty of data processing, model development and deployment, and O&M and management. The key software frameworks mostly adopt open-source frameworks such as TensorFlow, PyTorch, Paddle, MindSpore, and OneFlow. **At the development process level,** engineering focuses on the AI model development lifecycle process, pursues efficient and standardized continuous production, continuous delivery, and continuous deployment, and finally sends the best model to the application level to generate business value. For example, MLOps establishes a standardized model development, deployment, and O&M process to connect the model construction team, business team, and O&M team. **At the model management level,** with the gradual deepening of enterprise intelligentized applications, the types and number of models have increased significantly. Enterprises need to build a management mechanism for the model lifecycle and implement standardized management and O&M for model version history, performance, attributes, relevant data, and derived model files.

Automated machine learning (AutoML) technology is an important capability for improving engineering capabilities. Automated machine learning refers to the automation of all or part of the whole process of machine learning development and application. This can effectively reduce the challenges in the current stage of AI development, such as the high threshold for AI development and the lack of technical talents. This technology mainly includes automated data preprocessing, automated feature engineering, automated hyperparameter search, automated model network structure design, and automated model deployment. Technologies such as low-code development and pre-trained models are also closely related to automated machine learning and show a trend of integrated development. At present, leading Internet companies and innovative companies have begun to actively deploy AutoML technologies and tools. However, limited by the maturity of this technology, the application scenarios of AutoML still remain confined to stages of the development process (such as feature engineering) or some specific technical fields (such as speech recognition, object detection, or intelligent dialogue).

Technical demands for cloud-edge-terminal collaborative management have gradually risen in prominence, and the process of AI cloud migration continues to accelerate. With the deep integration of AI with various industries, AI edge and terminal devices will become increasingly widely used. At the same time, developers will also face the problems of complicated and difficult adaptation of edge devices and difficult O&M and management. On the one hand, the platform realizes model adaptation and deployment for edge devices through technologies such as model

compression and adaptive model generation. On the other hand, through the design and configuration of compilation optimization and intermediate representation, collaborative management and O&M can be implemented for cloud, edge, and terminal devices.

(iii) Safe and Trustworthy AI Technology Is Developing in the Direction of Integration

With the continuous attention of all sectors on the issue of trust in AI, safe and trustworthy AI technology has become a hot research field. The main focus of research is improvements to the **stability, explainability, privacy protections, and fairness of AI systems**. These technologies constitute the basic supporting capabilities of trustworthy AI.

The technical focus of AI system stability has gradually expanded from the digital domain to the physical domain. AI systems face unique attacks such as poisoning attacks, adversarial attacks, and backdoor attacks, which increase security challenges. These attack techniques can exist independently or simultaneously. For example, by printing adversarial sample eyeglasses, attackers can directly cause physical interference to facial recognition systems. Or, the attacker can paste an adversarial sample perturbation pattern on a street sign, which makes the autonomous driving system mistakenly recognize a "stop" sign as a "speed limit" sign. Stability testing technology centered around AI systems has also become the key. Huawei, Baidu, and other companies have launched related testing technologies based on fuzzy logic and are committed to exploring and improving the stability of AI systems.

AI explainability enhancement technology is still in its infancy, and various paths continue to be explored. Enhancing the explainability of AI systems has become a popular area of work, and the main paths include establishing appropriate visualization mechanisms to try to evaluate and interpret the intermediate states of the model; analyzing the impact of training data on the final converged AI model through the influence function; using methods to analyze which data features the AI model uses to make predictions; and investigating the explainability of black-box models by using simple interpretable models to locally approximate complex black-box models.

Privacy-preserving computation technology assists in safe and trustworthy AI data collaboration. AI systems must rely on a large amount of data, but the flow of data and the AI model itself may leak sensitive private data. The combination of AI and privacy-preserving computation technology can ensure the authenticity and credibility of the raw data from the data source. Using privacy-preserving computation technology, data is "available and invisible," forming a logically centralized view of physically dispersed multivariate data. This can ensure that sufficient and credible data

is available to AI models.

The key to improving the fairness of AI is to start from both data and technology. With the widespread application of AI systems, issues such as unfair decision-making behavior and discrimination against some groups have become more and more prominent. The main reasons for such decision-making biases are as follows: limited by data collection conditions, the weights of different groups in the data are unbalanced; when an AI model is trained on an unbalanced data set, the model decision-making becomes unfair. In order to ensure the fairness of decision-making in AI systems, at the data level, the main method is to construct completely heterogeneous data sets to minimize inherent discrimination and bias in the data; data sets are then checked periodically to ensure the high quality of the data. At the technology level, there are also algorithms that use fair decision-making quantitative indicators to reduce or eliminate decision-making bias and potential discrimination.

The systematic promotion of AI trustworthiness and safety technology will be an important trend. On the one hand, most of the current relevant research is carried out from a single dimension, such as stability, privacy, or fairness. Existing research work has shown that different requirements such as stability, fairness, and explainability are mutually synergistic or restrictive. If only one aspect of a requirement is considered, it may cause conflicts with other requirements. How to build a systemic research framework to maintain the optimal dynamic balance between different characteristic elements has become the key. On the other hand, it is necessary to carry out research on trustworthiness and safety from the system level. This problem is not only a problem at the level of AI algorithms. It also involves the entire system, such as the security problems of the operating system, software framework, third-party libraries, and hardware equipment used to run AI. It is necessary to build trustworthiness and safety for the full AI life cycle and chain.

III. The World Is Highly Attentive to AI Governance and AI Safety and Trustworthiness Has Become the Focus

The greater the space for AI development, the deeper its impact, and the more challenges it faces, the more important and urgent it is to govern it. At present, the world has formed a governance model of diverse entity participation and coordinated co-governance. Countries and organizations have introduced a series of governance principles, substantial progress has been made in the legislative process, and industry organizations and corporate entities are actively exploring trustworthy implementation practices.

(i) AI Risks Continue to Increase, and a Global Governance Mechanism Is Initially Established

1. In-Depth Empowerment of AI Raises Challenges

The risks and challenges posed by AI are manifold. In addition to the natural defects of AI technology itself, in contrast to purely technical risks, the origin of AI risks is the impact of the applications of AI systems on existing normative systems, ethics, and social order.

The inherent technical risks of AI continue to expand. AI technology with deep learning at its core is constantly exposing hidden risks arising from its own characteristics. First, deep learning models have the flaws of fragility and vulnerability, making it difficult to obtain sufficient confidence in the trustworthiness of AI systems. Second, black-box models have a high degree of complexity and uncertainty, which can easily lead to unpredictable risks. Third, the results generated by AI algorithms are overly dependent on training data. If there is bias and discrimination in the training data, it will lead to unfairness in intelligent decision-making.

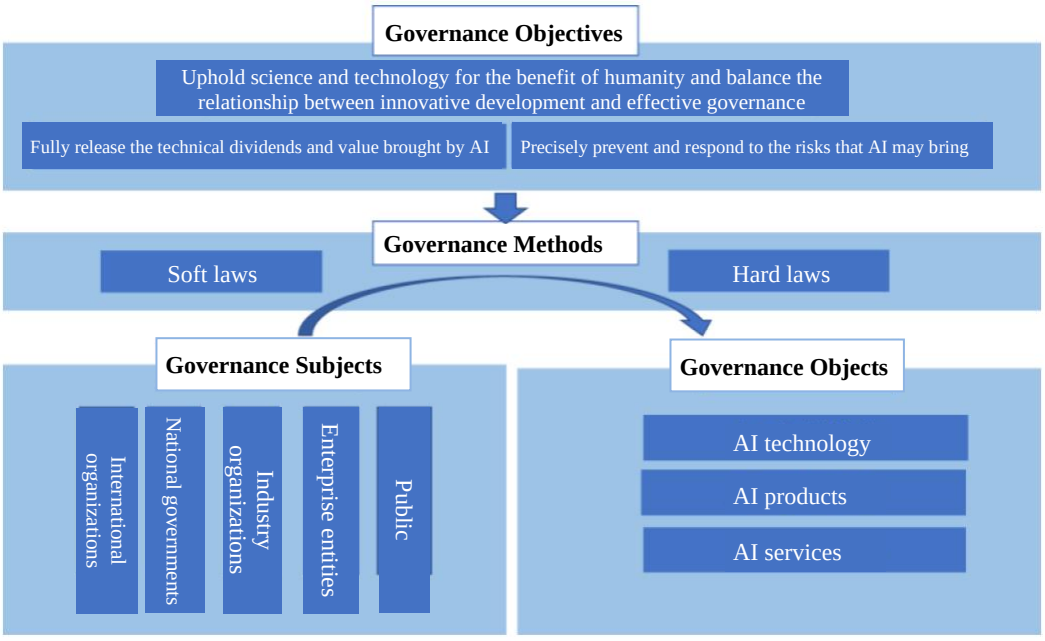
Challenges to existing legal and normative systems continue to expand. AI has impacted legal and normative systems in various ways: **In terms of entity qualification determination,** Saudi Arabia's grant of citizenship to the robot Sophia has sparked global controversy. In addition, questions such as whether AI can become the inventor of a patent have also arisen. For example, in July 2021, the Federal Court of Australia ruled that AI systems can be listed as inventors in patent applications, which is quite the opposite of the attitude of the United States and the United Kingdom. **In terms of privacy protection,** the development of AI is accompanied by the violation of personal privacy. The China Central Television (CCTV) "March 15" party was exposed, and a large number of companies have illegally collected customer face information to use for commercial purposes. **In terms of division of responsibilities,** in 2015, the first robotic surgery in the UK resulted in death, and Tesla's "loss of door control" incident cast doubt on the automated driving assistance system.

The impact on ethics and social order is growing increasingly serious. AI has the risk of impacting human rights. AI causes discrimination, proposes new rules for human behavior, and makes changes in the workforce. In August 2021, the Russian online payment service Xsolla used an algorithm to judge whether employees were "disengaged and inefficient" and fired 147 employees, which accounted for a third of the company's workforce. AI directly or indirectly harms humans and impacts social order. In November 2020, media reports claimed that an Iranian nuclear scientist was assassinated by a weapon controlled by "AI." In 2019, Amazon's smart speakers advised someone to commit suicide.

2. A Global Wave of AI Governance Has Arisen

At present, in the face of various risks and challenges arising from the in-depth

empowerment of AI, countries around the world are paying more and more attention to AI governance. AI governance is a complex systematic project. According to the *White Paper on Artificial Intelligence Governance*, an AI governance system consists of the joint participation and cooperation of multiple subjects, such as the government, industry organizations, enterprises, and the public. This forms a governance method combining "soft laws" such as ethical principles and "hard laws" such as laws and regulations, which aims to realize the overall goal and vision of science and technology for good and benefits for humanity and to promote the healthy and orderly development of AI. Figure 3 shows a schematic of AI governance mechanisms.



Data compiled by: China Academy of Information and Communications Technology (CAICT)

Figure 3 Schematic of AI Governance Mechanisms

The world's major economies focus on discussions of key issues in AI governance, and intergovernmental international organizations have become important voices. The United Nations, G20, OECD, and G7 have become important organizations guiding global AI governance. The relevant research results of the OECD have played an important role in promoting global AI governance. They are an important reference for G7 resolutions related to AI as well as the G20 AI Principles.

The **United Nations** (UN) actively promotes the process of AI ethical governance. The United Nations Educational, Scientific and Cultural Organization (UNESCO) issued its *Recommendation on the Ethics of Artificial Intelligence* on November 25, 2021. This is the world's first normative framework formulated for AI ethics and the broadest consensus reached at the governmental level in the world to date. At the same time, it gives each country the responsibility to apply the framework

at the corresponding level. The World Health Organization (WHO) published its first guideline on the use of AI in healthcare, *Ethics and Governance of Artificial Intelligence for Health*, on June 28, 2021. This ensures that AI technology can serve the public interest of all countries around the world.

In June 2019, the **Group of Twenty** (G20), with reference to the *OECD Principles on Artificial Intelligence*, approved the *G20 AI Principles*, which advocates the use and development of AI "with respect for legal principles, human rights, and democratic values." This became the first international intergovernmental consensus on AI governance and established a people-centered development concept. China supports strengthening dialogue centered around AI, implementing the G20 AI Principles, and promoting the healthy development of AI worldwide.

On May 22, 2019, the **Organization for Economic Cooperation and Development** (OECD) released the world's first intergovernmental policy guidelines on AI, forming the *OECD Principles on Artificial Intelligence*. This established five principles for the responsible management of trustworthy AI. The OECD established its AI Policy Observatory (OECD.AI) in February 2020 to share best examples of AI policy and practice, promote international cooperation, and help member states build trustworthy AI systems for the benefit of society as a whole.

The **Group of Seven** (G7) has launched an exploration on the consensus on AI governance among advanced economies around the world. The G7 summit in January 2021 indicates member states would collaborate on international AI standards; in September, at the G7 data protection and privacy authorities' meeting, they stated that data protection and privacy supervision will be the core work of AI governance in the future and urged industry to design AI products that meet data protection requirements.

(ii) AI Governance Has Entered a New Stage of Soft and Hard Law Coordination and Scenario Regulation

Since the publication of the *Asilomar AI Principles* in 2017, there has been a global upsurge in the exploration and formulation of ethical principles for AI. At present, the G20 AI Principles are widely recognized by the international community, intergovernmental organizations have become an important force in guiding the direction of AI governance, and countries around the world are accelerating the improvement of their relevant systems of rules for AI governance. The first draft of the EU's *Artificial Intelligence Act* in 2021 marked the acceleration of AI governance from the principled constraints of "soft law" to the more substantive regulation of "hard law." At the same time, with the deepening of the integration between AI and the real economy, AI governance has become increasingly focused on specific scenarios.

1. Process of AI Governance Materialization Has Accelerated

At present, the focus of AI governance in different countries is different, but on the whole, it shows a trend of accelerated evolution. That is to say, from the early stage of building a system of social norms system guided by "soft law," it is beginning to move towards a risk prevention and control system guaranteed by "hard law."

The EU is steadily advancing from ethics to regulation and desires to take the lead in global AI regulation rules. On April 21, 2021, the European Union published the draft of the *Artificial Intelligence Act*. This is the first law in the world to systematically regulate AI. It refines the four-level AI risk framework, focuses on regulating high-risk systems, and proposes relatively complete regulatory support measures. This is another important move by the EU following the publication of the *Ethics Guidelines for Trustworthy AI* (2018) and the *White Paper on Artificial Intelligence: a European approach to excellence and trust* (2020). It marks the shift of global AI governance from soft constraints such as ethical principles to a stage of comprehensive and operable legal regulations.

The United States emphasizes prudential supervision to promote innovative development. The 2019 executive order *Maintaining American Leadership in Artificial Intelligence* established that the overall tone of the United States in AI governance was centered on strengthening its global leadership. The United States proposed the *Algorithmic Accountability Act* in 2019, requiring impact assessments on "high-risk" automated decision-making systems. In 2020, the U.S. Senate introduced the *National Biometric Information Privacy Act*, which provides privacy protection for AI-enabled biometric identification on the basis of personal privacy data protection. In May 2021, the U.S. *Algorithmic Justice and Online Platform Transparency Act* proposed obligations and requirements on algorithm transparency in terms of three entities: users, regulators, and the public. In July 2021, the U.S. Government Accountability Office released an AI accountability framework to ensure the fairness, reliability, traceability, and governance of AI systems.

In China, both soft and hard laws are taken into consideration, and both are used to promote AI governance. At the level of principles and ethics, the National New Generation Artificial Intelligence Governance Specialist Committee, after releasing the *Governance Principles for New Generation Artificial Intelligence: Developing Responsible Artificial Intelligence* in June 2019, released the *Ethical Norms for New Generation Artificial Intelligence*, which aims to integrate ethics into the full AI lifecycle and actively guide society as a whole to carry out AI R&D and application activities responsibly. **In terms of legal process,** China has not yet issued a unified law related to AI, but the *Personal Information Protection Law* officially implemented in November 2021, together with the *Cybersecurity Law* and the *Data Security Law*, form

a solid legal system for governing the underlying elements of AI. In addition, active explorations are being made at the local level. Shenzhen issued the *Shenzhen Special Economic Zone Artificial Intelligence Industry Promotion Regulations (Draft)* in July 2021 to assist with the healthy development of the AI industry.

At the same time, the United Kingdom, France, Japan, South Korea, and other countries have also carried out work related to AI governance. The United Kingdom emphasizes the development of AI norms and promotes AI education and talent training. *Artificial intelligence: opportunities and implications for the future of decision making* (2016), *AI in the UK: ready, willing, and able?* (2018), the *Emerging Technology Charter* (2021), and many other documents and reports call for the establishment of an AI code and ethical framework at the national level. France has deepened its understanding of the ethical issues of AI through expert seminars and academic debates. Japan, South Korea, and other countries are attentive to the ethics of AI from the perspective of the development of intelligentizatized transformation of manufacturing and the application of emerging technologies.

2. In Typical Scenario-Based Governance, Each Accelerates Implementation with Its Own Focus

The complexity of AI governance is also reflected in the diversification and differentiation of its application scenarios. In different scenarios, the application depth and impact of AI technology vary. The governance of typical scenarios has become the focus of work in various countries, especially in areas such as autonomous driving, smart healthcare, and facial recognition.

In the field of autonomous driving, Germany has taken the lead in formulating ethical guidelines and framework laws, and various countries have stepped up the deployment of graded and categorized (分级分类) supervision. Germany introduced its *Ethics Code for Automated and Connected Driving* in 2017 and passed the draft *Act on Autonomous Driving* in May 2021. In 2021, the UK discussed and amended the *Highways Act* to introduce new provisions for the safe use of autonomous vehicles on motorways. In China, in May 2021, the Cyberspace Administration of China, together with relevant departments, drafted *Several Provisions on the Management of Automobile Data Security (Draft for Comment)* to solicit public opinions.

In the field of smart healthcare, ethical principles have gradually developed, and the regulatory level focuses on regulating medical device access. Based on the 2019 *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) (Discussion Paper)*, the U.S. FDA released the *Artificial Intelligence/Machine Learning-*

Based Software as a Medical Device Action Plan in January 2021 to deploy regulatory initiatives for AI-based medical device software. The European Union introduced the Medical Device Regulation (MDR), requiring new medical devices to apply for a certificate of compliance starting May 2021. In June 2021, China issued the *Guiding Principles for the Registration and Review of Artificial Intelligence Medical Devices (Draft for Feedback)* (人工智能医疗器械注册审查指导原则 (征求意见稿)) and promoted the orderly development of the AI medical device industry.

In the field of facial recognition, countries around the world have entered the era of strong supervision of privacy protection and information and data security. The European Union included facial recognition in the high-risk categorization level in the draft *Artificial Intelligence Act* introduced in April 2021. In October, the European Parliament voted to pass a resolution calling for a complete ban on large-scale surveillance based on AI biometric technology. China implemented the *Personal Information Protection Law* in November 2021, and the judicial interpretation related to facial recognition issued by the Supreme People's Court in August specifically regulates the processing of face information. U.S. legislation at the state and local levels prohibits government agencies from using facial recognition technology in public places. The United Kingdom released the *Emerging Technology Charter* in September 2021, pointing out that technologies such as facial recognition must be used legally and ethically.

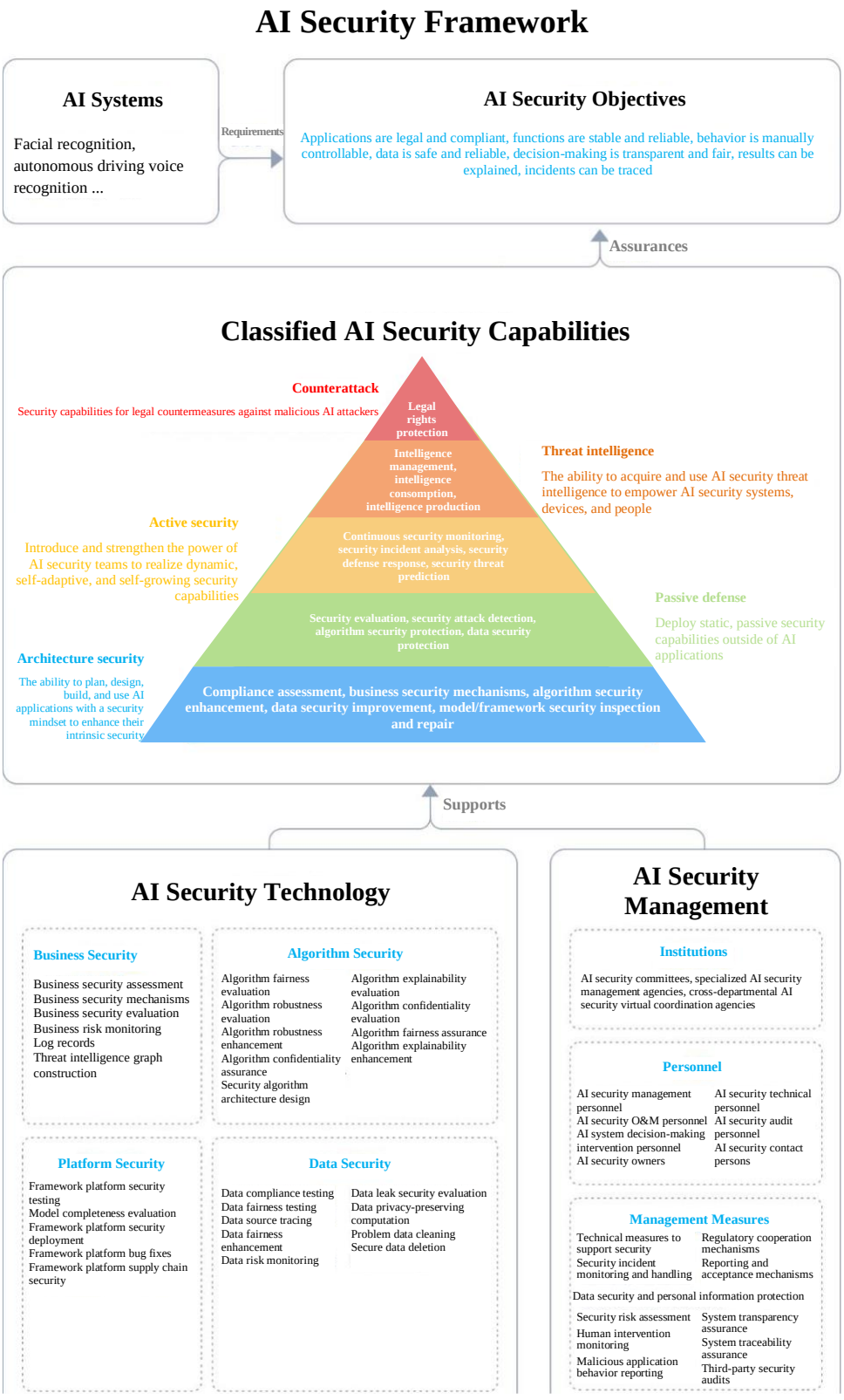
(iii) AI Security Frameworks Have Become a Key Guideline for Effective Risk Prevention

In order to effectively prevent the security risks brought about by the application of AI technology and to ensure the safety of AI systems that are related to national security, economic lifelines, and social stability, it is urgently needed to propose a security system for AI systems and provide effective guidance for industry in order to gradually improve AI security capabilities. AI security frameworks are based on the needs of AI security protection and organically integrate AI security technology systems and AI security management systems. The overall system design and planning of AI security constructed is of great significance to maintaining national AI security and cybersecurity.

1. AI Security Frameworks Are Gradually Taking Shape

An AI security framework needs to include four dimensions: security goals, security capabilities, security technologies, and security management, as shown in Figure 4. These four protection dimensions guide enterprises to build AI security protection systems based on a top-down, layer-by-layer approach. In this effort, setting reasonable security goals is the starting point and basis for ensuring the

security of AI applications, security capabilities provide effective assurance for the achievement of security goals, and security technologies and security management are the pillars and embodiments of security capabilities.



Source: CAICT

Figure 4 AI Security Framework

2. Categorization and Grading Have Become a New Direction of Framework Construction

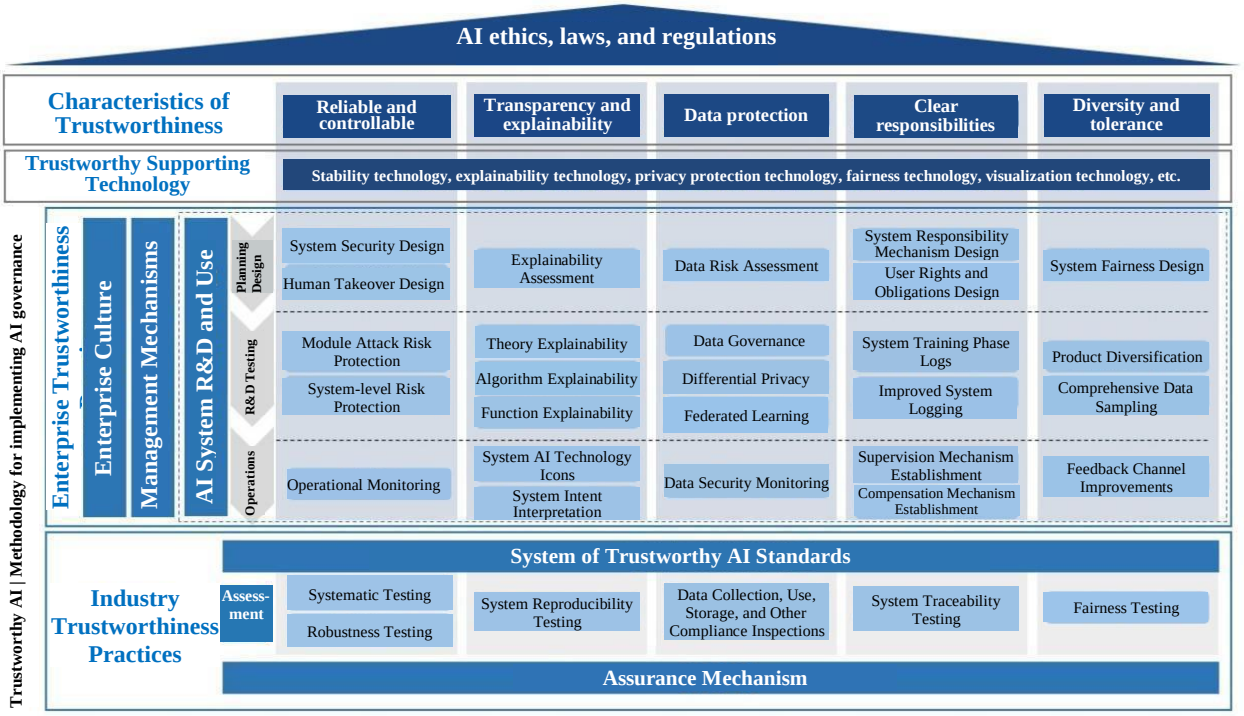
Categorization and grading have become a new trend in global AI governance. The EU draft *Artificial Intelligence Act*, the U.S. *Algorithm Accountability Act*, Canada's *Directive on Automated Decision-Making*, and China's *Guiding Opinions on Strengthening the Overall Governance of Internet Information Service Algorithms*, and other laws, regulations, and policy documents of major countries around the world have proposed to establish requirements for the categorized and graded management of AI systems and algorithms. However, the above-mentioned laws and regulations only propose categorization and graded management requirements or describe categorization methods by enumerating typical AI systems. They lack categorization principles, there is no grading method or process that can be followed, and they cannot be applied to rapidly emerging new AI applications. It is an urgent need to propose an AI categorization and grading system, clarify the categorization and grading principles, and facilitate categorization and grading elements and methods for actual operations.

According to the idea of categorized management and graded protection, this White Paper proposes the following recommendations for AI categorization and grading. According to their degree of autonomy, AI systems are divided into three categories: assisted human intelligence systems, human-machine hybrid intelligent systems, and fully autonomous intelligent systems. According to their importance and degree of harm, AI systems are divided into three levels: low-risk intelligent systems, high-risk intelligent systems, and ultra-high-risk intelligent systems. Each of these types of AI systems can be further divided into three levels.

(iv) Trustworthy AI Has Become an Important Methodology for Implementing Governance Requirements

According to the *White Paper on Trustworthy Artificial Intelligence*,[4] facing the global anxiety caused by a lack of trust in AI, the development of trustworthy AI has become a global consensus. **Trustworthy AI is a set of methodologies for implementing AI governance requirements from the industrial dimension and a bridge between AI governance and industrial practices.** Figure 5 shows the overall framework of trustworthy AI.

[4] Translator's note: For an English translation of the CAICT-JD *White Paper on Trustworthy Artificial Intelligence*, see: https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/.



Source: CAICT

Figure 5 Overall Framework of Trustworthy AI

1. The Trustworthy Concept Is Gradually Penetrating the Whole AI Life Cycle

Trustworthy AI was proposed by academia, actively researched by many sectors, and then started to be implemented in practice by industry circles. Its significance is gradually becoming richer and evolving. The definition of trustworthy AI is no longer limited to the statuses of AI technology, products, and services themselves, but instead has gradually expanded to a set of systematic methodologies, involving all aspects involved in building "trustworthy" AI. This includes internal management, R&D, operations, and other aspects of enterprises as well as industry-related work to translate relevant abstract requirements into specific capability requirements for practices, thereby enhancing society's level of trust in AI.

2. Enterprises Have Become a Major Force in Practicing Trustworthy AI

As the front line of AI technology R&D and innovative application, enterprises need to face the challenge of AI trust, actively carry out self-discipline and self-governance work, and give full play to the initiative of enterprises in implementing the trustworthiness requirements of AI technology, products, and services. Since 2018, many domestic and foreign companies such as Google, Microsoft, IBM, Megvii, and Tencent have launched corporate AI governance guidelines and formed corresponding departments and agencies to promote the implementation of

governance responsibilities. In addition, enterprises are also actively exploring AI governance models with the practice of trustworthiness as the core concept. IBM, Microsoft, Huawei, JD, and other Chinese and foreign enterprises have released a number of AI trustworthiness tools in order to help AI products improve security, robustness, explainability, fairness, and other trustworthy capabilities in the process of R&D and to bring developers together to promote the concept of trustworthiness through the open-source ecosystem.

3. Industry Organizations Promote the Creation of a Safe and Trustworthy Ecosystem for AI

The realization of trustworthy AI is not only accomplished by the unilateral practice and efforts of enterprises but also requires the participation and coordination of multiple parties. Ultimately, a healthy ecosystem of mutual influence, mutual support, and interdependence must be formed. **At the level of standards formulation**, since 2017, ISO/IEC, IEEE, SAC/TC 28/SC 42, and other Chinese and foreign standards organizations have taken the lead in laying out universal standards for trustworthy AI. In April 2021, the Chinese national standard *Information Security Technology Facial Recognition Data Security Requirements* was opened to the public for comments. **At the level of industry self-discipline**, China's Artificial Intelligence Industry Alliance released the *Joint Pledge on Artificial Intelligence Industry Self-Discipline* in 2019. Subsequently, in 2020, they released the *Guidelines for Trustworthy AI Operations* and announced the first batch of commercial AI system trustworthiness evaluation results, involving 16 AI systems from 11 companies. This provided an important model selection reference for users. At present, the *Administrative Guidelines for Trustworthy Artificial Intelligence R&D* and other documents are being jointly compiled with industry circles in order to promote the safety and trustworthiness of AI R&D at the source.

IV. Summary and Outlook

Chinese AI technology and industry have made great progress in development. We believe that during the "14th Five-Year Plan" period [2021-2025], AI technology innovation will be further accelerated, the scale of the industry will continue to expand, and a number of high-quality enterprises and industrial clusters with great development potential will emerge, becoming an important engine in leading the high-quality development of the economy.

Pursuing technological innovation, focusing on engineering practices, and ensuring trustworthiness and safety have gradually become important directions for the future development of AI. Looking back at the development of AI over the past ten years, it is not difficult to find that technological innovation and engineering practices

complement each other. Breakthroughs in algorithms and computing power have driven the development of tool systems, and the maturity of tools has further supported the application of technology. At present, AI is already widely used in all aspects of people's daily work and life, and the demand for its safety, trustworthiness, and quality has risen to an unprecedented level. Promoting the reliable and controllable development of AI has become a global consensus. Standing at the outset of the "14th Five-Year Plan," we look forward to the continuous improvement in AI technology and the vigorous and healthy development of the AI industry and applications over the next five years.

First, while constantly exploring new technologies, we must pay more attention to releasing technological dividends through engineering methods and ensure safety and trustworthiness. The key factor that determines whether AI companies can quickly empower all industries and sectors and respond to diverse needs is the engineering capabilities of enterprises. At the same time, the demand for safe and trustworthy technology is becoming more and more important. Currently, a large number of companies engaged in privacy-preserving computation technology have emerged, centered around data protection. In the future, technologies centered around AI stability and fairness will also form an important force.

Second, in the process of industrial intelligentization, the level of participation of traditional industries will be more and more in-depth, and they will even lead the development process of the entire industry. The focus of industrial development has already begun to shift from "AI+" to "+AI." With the improvement of the digitalization process of traditional industries, vast quantities of data and rich application scenarios will be provided, opening up new space for the application of AI. In these traditional industries and fields, institutions with higher AI penetration rates will output AI-related solutions to other institutions throughout their fields.

Third, AI governance will become more and more critical. It is related to the sustainable and healthy development of AI, and the coordination of governance and development has become a necessity. Governance work is not only practically related to the day-to-day application of AI, but has also become an important topic of international competition and cooperation. In the face of different cultural backgrounds and levels of development in different countries and regions around the world, how to effectively carry out AI governance practices is an important challenge. The Chinese government, industry organizations, and enterprises have taken the lead in starting to explore AI governance, integrating the concept of safety and trustworthiness into the full AI lifecycle. In the future, more practice paradigms with also emerge.