



世界互联网大会
World Internet
Conference

Developing Responsible Generative Artificial Intelligence Research Report and Consensus

World Internet Conference Working Group on Artificial Intelligence
November 2023

WORLD INTERNET CONFERENCE WORKING GROUP ON ARTIFICIAL INTELLIGENCE

listed in alphabetical order by units

-
- **360 Security Technology Inc.**
Zhaohui Liu, Chuang Liu, Yiyun Zhen
 - **Alibaba Cloud**
Ran Zhu, Fei Mu, Yudong Yang
 - **Ant Group Co., Ltd.**
Guanchen Lin, Zujie Wen, Liang Li
 - **Arab ICT Organization**
Mohamed Ben Amor
 - **Arm Technology (China) Co., Ltd.**
Dawei Wu, Junchao Wang, Tong Wu
 - **Beihang University**
Jingwu Zhao
 - **Beijing Academy of Artificial Intelligence**
Yonghua Lin, Xianhao Ni
 - **Beijing Baidu Netcom Science Technology Co., Ltd.**
Yanjun Ma, Yanli Liu, Yujie Wang
 - **Beijing Sankuai Technology Co., Ltd.**
Jingang Wang, Zhengyu Chen
 - **Beijing SenseTime Technology Development Co., Ltd.**
Feng Tian, Zhengkun Hu
 - **Beijing Zhipu Huazhang Technology Co., Ltd.**
Debing Liu, Miao Zhang
 - **British Standards Institution**
Chris Brown
 - **Build n Blaze; ITU-T Focus Group on metaverse**
Radia Funna
 - **CEN-CENELEC JTC 13 Cybersecurity and Data Protection**
Walter Fumy
 - **China Academy of Information and Communications Technology**
Kai Wei, Yuntao Wang
 - **China University of Political Science and Law**
Linghan Zhang
 - **China Renaissance**
Chuan Qin, Lixing Wang, Yuping Zhao
 - **Cisco Systems, Inc.**
Shengli Hou, Hang Wei
 - **FUXI INSTITUTION**
Xiaodong Li, Wei Fu, Xiaobo Yang
 - **Guangzhou Dongyue Information Technology Co., Ltd.**
Guanjun Jiang, Yuan Zhang
 - **Henley Business School, UK**
Yinshan Tang
 - **Huawei Cloud Computing Technology Co., Ltd.**
Peng You, Yin Li, Qian Zhang
-

WORLD INTERNET CONFERENCE WORKING GROUP ON ARTIFICIAL INTELLIGENCE

listed in alphabetical order by units

- **HUAWEI Technologies Co., Ltd.**
Zhenjun Wang, Jiaqi Liu
 - **IBM (China) Company Limited**
Dong Xie, Fanjing Meng, Haixu Cheng
 - **Iflytek Co., Ltd.**
Shijin Wang, Sen Li, Qiaolin Dong
 - **Institute of Automation, Chinese Academy of Sciences**
Yi Zeng
 - **Institute of Law, Chinese Academy of Social Sciences**
Hui Zhou
 - **Institute of Philosophy, Chinese Academy of Social Sciences**
Weiwen Duan
 - **Intel China Ltd.**
Haining Wang, Mario Romao
 - **Kompetus Management Consulting GmbH, Germany**
Felix Richter
 - **ITU-T Focus Group on Autonomous Networks**
Vishnu Ram O.V.
 - **PCI Technology Group Co., Ltd.**
Zhiwen Zhou, Wei Qin
 - **NetEase (Hangzhou) Technology Co., Ltd**
Tangjie Lyu, Mengjie Ma
 - **Nokia Shanghai Bell Co., Ltd.**
Chenhui Ye, Zhuoran Xiao, Jiang Chang
 - **ITU Telecommunication Standardization Bureau**
Xiaoya Yang
 - **Sina Corporation**
Wei Wang, Junlin Zhang, Xueting Wang
 - **Tencent**
Jianfeng Cao, Mengyin Wang, Jun Yuan
 - **University of Muenster, Germany**
Bernd Holznagel
 - **University of International Business and Economics**
Xin Zhang
 - **Tsinghua University and the Beijing National Research Center for Information Science and Technology**
Xinyi Gu
 - **Yinxiang Biji**
Yi Tang, Feng Yue, Ying Bu
 - **World Internet Conference**
Hao Liang, Xueli Zhang, Yanrong Kang
 - **ZTE Corporation**
Wei Meng, Liya Yuan
 - **Zhejiang Lab**
Feng Lin, Zhe Liu
-

Editors

China Academy of Information and Communications Technology
Naying Hu, Lin Shi, Shuo Liu, Hao Zou, Xiaotong Sun, Sumin Guo, Lingyun Zhu

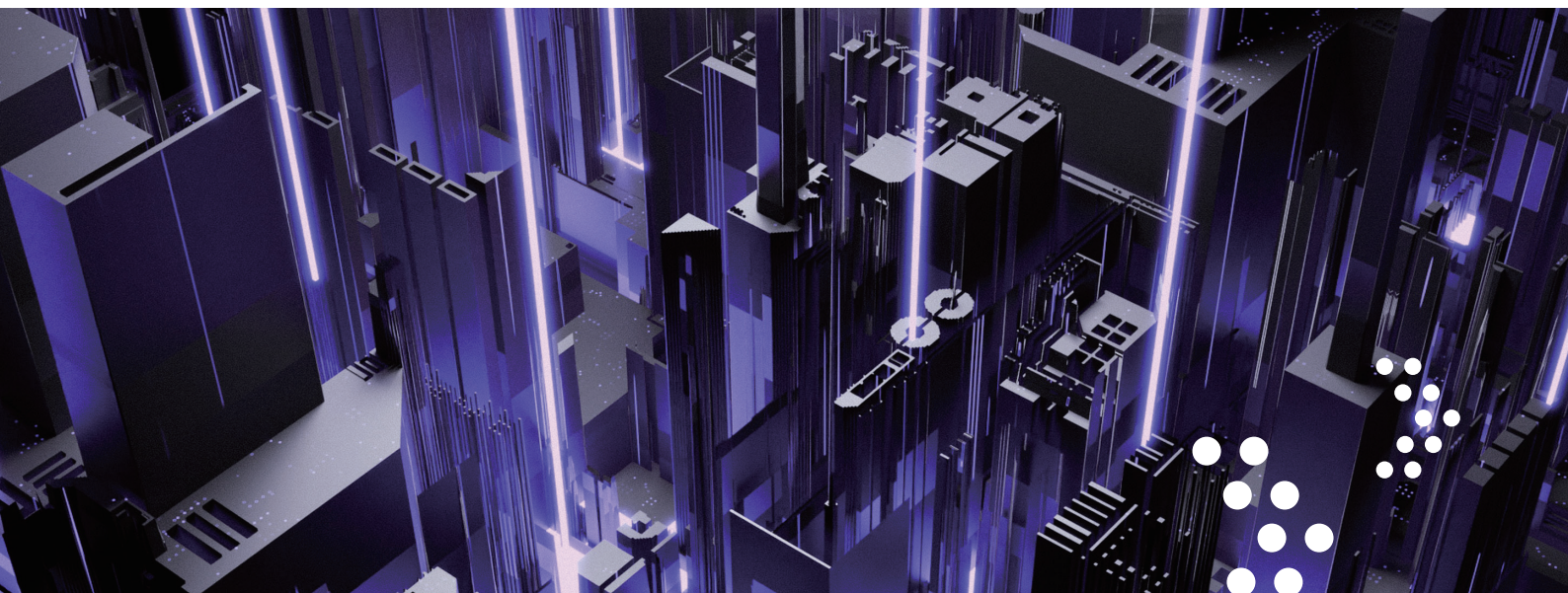
Contact us: research@wicinternet.org

ABBREVIATION TABLE

3D	Three Dimensional
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
BLIP	Bootstrapping Language-Image Pre-training
BLOOM	BigScience Large Open-science Open-access Multilingual Language Model
Chat GLM	Chat Generative Language Model
CPU	Central Processing Unit
DALL-E	Dali WALL-E
Emu	Large Multimodal model of BAAI
FPGA	Field-Programmable Gate Array
GenAI	Generative Artificial Intelligence
GDP	Gross Domestic Product
GMV	Gross Merchandise Volume
GPT	Generative Pretrained Transformer
GPU	Graphics Processing Unit
InternLM	Intern Language Model
LLaMA	Large Language Model Meta AI
MTP	Massive Text Pairs
NLP	Natural Language Processing
NPU	Neural Processing Unit
PaLM	Pathways Language Model
Qwen	Qian Wen
SaaS	Software as a Service
SDK	Software Development Kit
TPU	Tensor Processing Unit
VLA	Vision Language Action
VLM	Vision Language Model

CONTENT

I. Overview	01
II. Global Development Trends of GenAI Technologies	02
III. Opportunities Brought by GenAI	06
IV. Challenges Brought by GenAI	10
V. Global Efforts to Develop Responsible GenAI	13
VI. Consensus on Developing Responsible GenAI	18
Attachment: Exploration of Industrial Applications for Developing Responsible Generative Artificial Intelligence	22



01/ OVERVIEW

In recent years, Generative Artificial Intelligence (GenAI) has made continuous breakthroughs, showcasing both powerful creative generation abilities and the possible emergence of “wisdom”. It’s advances in understanding and generation of text, code, image, audio, video, etc., which is expected to greatly enhance social productivity, accelerate the digitalization process of industries, and promote the comprehensive development of human society towards a new stage of intelligence.

The more than 60-year history of artificial intelligence (AI) is one where technological breakthroughs create significant opportunities for development while ushering in corresponding challenges. Coordinating the development and governance of AI is becoming a global consensus.

Since 2016, numerous international organizations, countries, regions, and industries worldwide have been actively exploring the development and governance of AI. A series of consensus, principles, governance requirements and practice paradigms have been formed globally. Considering that AI is still developing rapidly, sustained efforts need to be promoted continuously.

GenAI’s rapid rate of evolution, extensive scope of empowerment, and unprecedented depth of impact is an increasingly urgent call for sustained efforts to promote its development in a responsible manner. This issue has grown in importance for the development of AI, and perhaps even human civilization. For this reason, the World Internet Conference has established its Working Group on Artificial Intelligence to gather knowledge and perspectives from various stakeholders to reach a global consensus regarding the collaborative development and governance of GenAI for the improvement and general welfare of all humanity.

02/ Global Development Trends of GenAI Technologies

(1) Breakthroughs in AI Driven by Three Elements “model, data, computing power”

Continued advancements in AI is the result of the development of three fundamental components: **model, data, and computing power.**

In model dimension, improvements in model structure and scale are crucial factors in driving the advancement of GenAI. Regarding the structure of the model, various technologies such as the attention mechanism, autoregressive model, and diffusion model are consistently improving the iteration process. Notably, the foundation model, originating with Transformers

has emerged as the dominant technological approach for generating models; enabling the understanding and creation of text, images, audio, video, and other types of content. Large language models such as ChatGPT and ERNIE Bot, visual generation models like Stable Diffusion, DALL-E2, DALL-E3, as well as multi-modal models like GPT-4, BLIP-2, and Emu have emerged. **Regarding the scale of model parameters,** the aforementioned novel model architecture enables an increase in parameter scale, thereby greatly enhancing the capacity of the model. The GPT-3 series model, introduced in 2020, had a scale of 175 billion. The GPT-1 model, released in 2018, had a scale of 1.1 billion. Consequently, the GPT-3 model has demonstrated significant advancements in the field of complex natural language processing. Furthermore, by integrating external search, data processing, and other functions with the foundation model capabilities, the plug-in mechanisms derived from the base model can further enhance the model's functionality and broaden its scope of applica-

tion. Prominent corporations, such as OpenAI, 360, Baidu, HUAWEI, iFLYTEK, and others, have implemented model plug-ins. ERNIE Bot, for example, possesses online search, interaction, and other plug-ins that facilitate the expansion and modification of the model. This enhances its functionality and allows it to accommodate a wide range of scenarios.

In data dimension, improvement of data quality, diversity, and scope, among other factors, is crucial for the advancement of GenAI capabilities. The Pile dataset, which is extensively used for pre-training in large models, is of high quality and mainly consists of over 20 datasets obtained from sources such as Wikipedia, books, journals, Reddit links, Common Crawl, etc¹. The Beijing Academy of Artificial Intelligence has released a massive collection of text pairs (MTP) in the form of a 300 million-pair text. This collection includes encompassing search content, community Q&A, encyclopedia knowledge, scientific literature. Anthropic, Stanford University, and Hugging Face have also released fine-tuning datasets that cover a wide range of instruction categories. These datasets aim to enhance the model's understanding and the ability to follow human instructions, thereby improving its controllability. **Furthermore, synthetic data can serve as a valuable source of high-quality data.** The ability of GenAI to generate large-scale synthetic data represents a promising approach to addressing the potential future problem of depleting high-quality training data. According to Gartner, by 2024, 60% of the data used for AI development and analysis will be synthetic data. By 2030, synthetic data will have replaced significant amount of real data and become the primary data source for AI models².

In computing power dimension, the advancement of computing infrastructure has enabled the exponential growth of GenAI. AI Chips provide fundamental support for computing power. For AI computing, various technological paths, including GPU, FPGA, NPU, TPU, and others, are perpetually investigated and optimized, providing a solid foundation for model training and inference. **Deep learning frameworks enhance the power efficiency of chip computing.** Firstly, it provides large-scale distributed training and inference technology with high performance, thereby effectively mitigating the challenges of lengthy model training times and high demand for inference computing power. Secondly, it optimizes and maximizes hardware performance and computational efficiency through adaptation and optimization of underlying chips. **Cloud-edge-end Computing satisfies various GenAI applications.** In terms of cloud-side computing, it processes robust computational and storage capabilities to support the successful training of large models and executing inference tasks for high-throughput applications. In terms of edge-side computing, which supplements cloud-side computing power, it can localize and preprocess massive and complex data in real time. Then direct these data to large models for rapid response and decision-making. In terms of terminal-side computing, conducting data computation and analysis on the terminal side reduces data transmission and processing latency, thereby enhancing the real-time performance of intelligent applications.

(2) Thriving GenAI Ecosystem Driven by Openness

¹ Leo Gao, Stella Biderman, et al. The Pile: An 800GB Dataset of Diverse Text for Language Modeling, Dec. 2020, p1.

² Gartner, 'Maverick Research : Forget About Your Real Data – Synthetic Data Is the Future of AI' (24 June 2023), <<https://www.gartner.com/en/documents/4002912>> accessed 19 September 2023.

Open-source models facilitate the development and popularization of technology. Represented by LLaMA 2, BLOOM, Chat GLM, Baichuan, Aquila, InternLM, Qwen, and others, open-source models are driving the iterative optimization of models and technological innovation. **In terms of model optimization and iteration,** the thriving of open-source models has expanded the options available to enterprise for both foundational models and fine-tuning models. Presently, a significant number of startups are introducing new products utilizing open-source models such as LLaMA 2 and Stable Diffusion. **In terms of research and development (R&D) threshold reduction,** application of open-source models offers benefits including the ability to avoid substantial initial investments, maintain full control over private data, and optimize self-imitation. Developers can rapidly build vertical task models with professional domain knowledge based on open-source models, significantly reducing the computational power, data, and time costs required for model development and application. For example, according to the open-source platform Github, the ChatGLM open-source model developed based on “Zhipu AI” has considerably lowered the research and development threshold, with 11 models standing out, covering multiple fields such as healthcare, law, finance, and education³.

Open interfaces provide convenience for developers. Beyond open-sourcing models, providing accessible and user-friendly APIs and SDKs is a pivotal component in fostering a thriving AI ecosystem. **On the one hand, open interfaces streamline development procedures and increase productivity.** With the help of open interfaces, developers are no longer need to develop algorithms or models start from scratch,

which greatly reduces development time and effort. An illustration of this can be seen in the minimal amount of Python code required to execute complex functions, such as code generation, dialogue agents, language translation, and auxiliary learning, using the GPT-3.5-Turbo model API’s open interface. **On the other hand, open interfaces enhance the application scenarios of models.** Open interfaces can assist developers in accessing model capabilities more conveniently, thereby enabling a wider range of application scenarios. ERNIE Bot, provided by Baidu, can be applied to various scenarios such as search, recommendation, dialogue, and others to enhance effectiveness and improve user experiences.

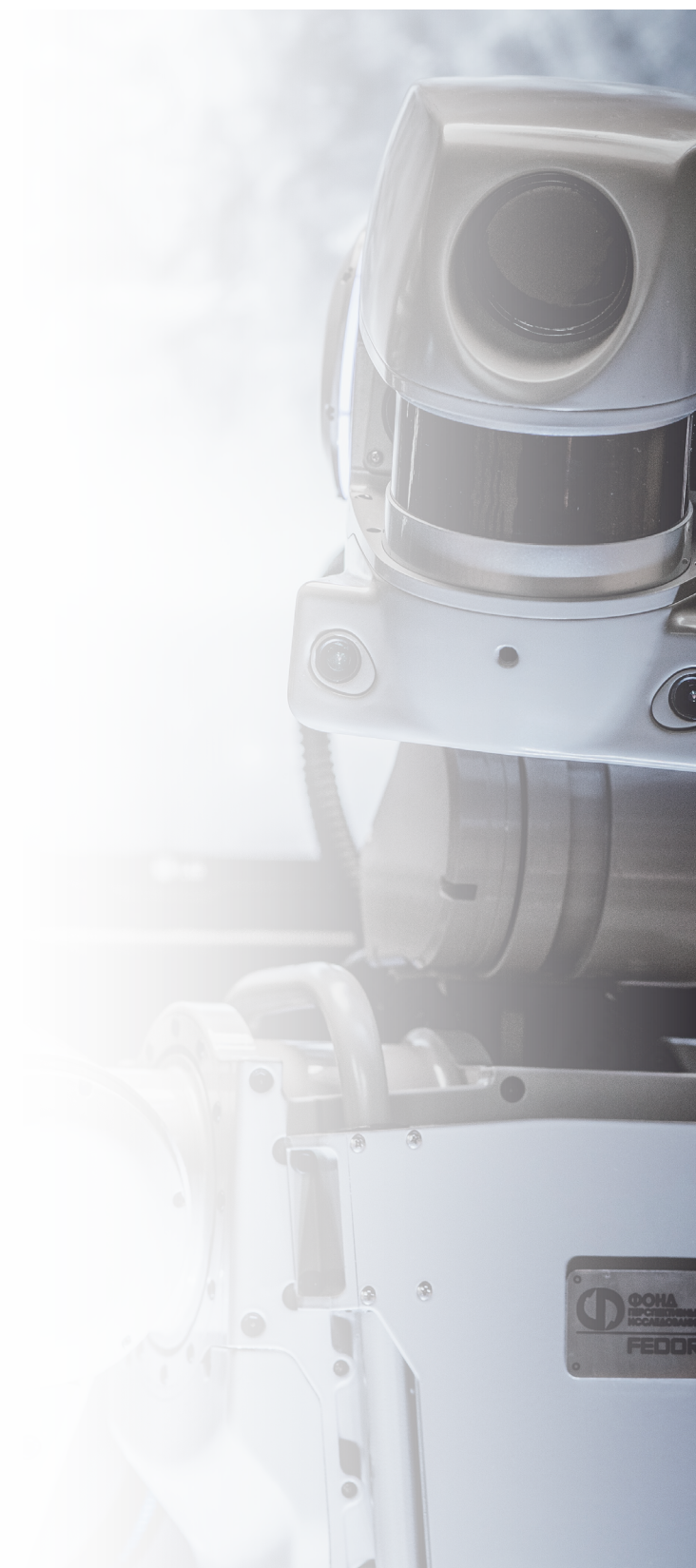
Developer communities can advance the diffusion of technology. By offering free computing power, course materials, public datasets, model kits, and other toolsets, developer communities empower and cultivate individuals with model development skills. As a result, they play a positive role in promoting technical communication and advancement in the field of AI. Hugging Face, for example, offers a one-click pre-training model calling function. It provides massive pre-training models, user-friendly APIs, abundant documentations, and active community forums to assist in technology diffusion. Huawei Cloud’s “AI Gallery” has developed a one-stop AI community service platform to assist enterprises and developers in quickly creating model applications. Alibaba’s ModelScope community provides developers with a large number of pre-trained models, as well as the ability to experience of these models online. This allows developers to quickly access different models without the need for coding. FlagOpen unifies the tools, evaluations, algorithms, models, and other compo-

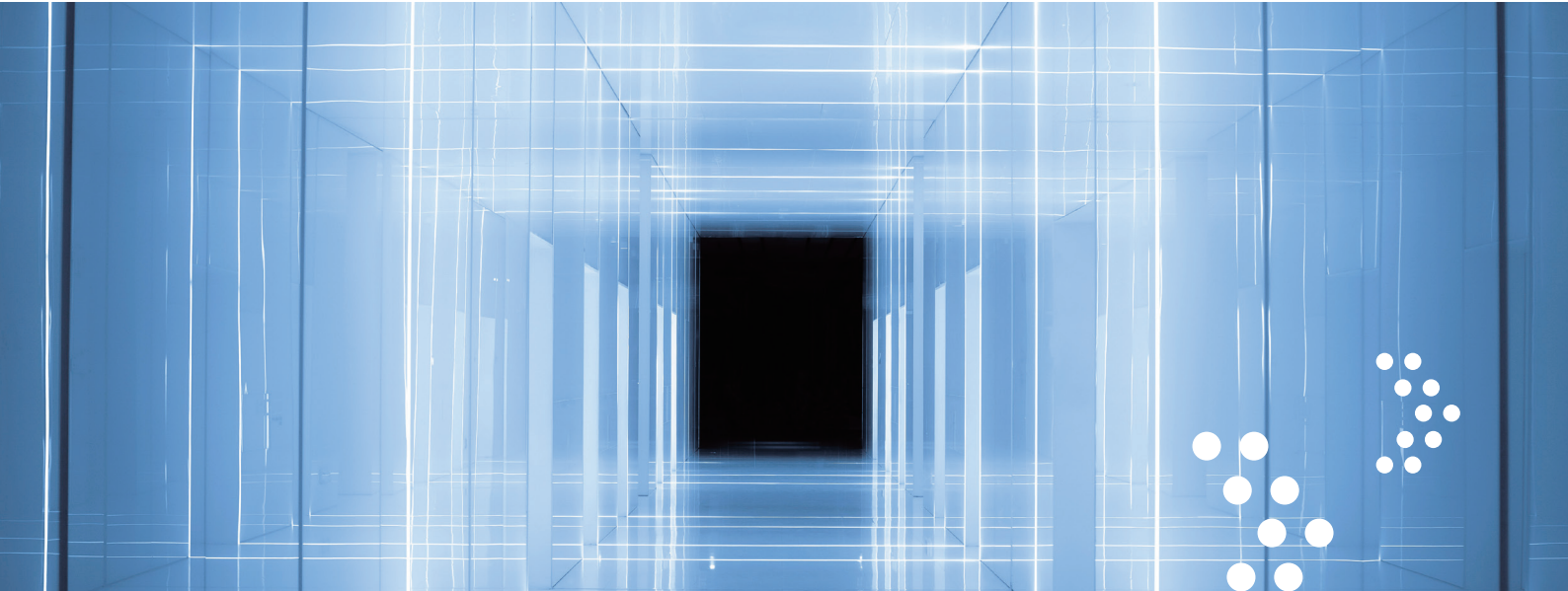
³ Github, ‘Awesome-Chinese-LLM’, <<https://github.com/HqWu-HITCS/Awesome-Chinese-LLM>> accessed 30 October 2023.

nents of complex models into an open-source “Linux” technology system. The Baidu “Paddle-Paddle” AI Studio facilitates the exploration of large models by providing integrated big model development system, open data, open-source algorithms, and free computation.

(3) The Development of GenAI Highlights the Emergence of AGI

Breakthroughs in GenAI has accelerated the pace of artificial general intelligence (AGI). Not only can GenAI handle tasks involving a single type of data, it can also connect and integrate different types of data, moving towards a multi-modal direction. **Breakthroughs in multimodal generative models have stimulated significant improvements in anthropomorphism and generalization of machine intelligence.** AI Agents, coupled with advancements in multimodal generative model technologies, will enhance their understanding and processing of complex real-world situations, thereby offering more precise and personalized services for humans. **The combination of multimodal generative models with AI Agent will bring more possibilities.** Embodied AI combines multimodal generative models with robotics to perceive the complex world by simulating human learning. This integration of hardware and software, combining multiple senses and cognitive abilities, will assist humans in completing various tasks. For instance, Google has released the Robotic Transformer 2 (RT2), which serves as a Vision-Language-Action (VLA) model. It integrates Vision-Language Model (VLM) pre-training with robotic data, allowing for direct control of robots to carry out various tasks in the real world.





03/ Opportunities Brought by GenAI

(1) Promote Economic Development

GenAI has great potential to significantly increase productivity and boost global economic growth. McKinsey estimated in June 2023 that GenAI could potentially add USD 2.6 trillion to USD 4.4 trillion in value to the global economy annually⁴. Goldman Sachs Research suggests that breakthroughs in GenAI could raise global GDP by 7% over the next decade⁵. At present, the productivity enhancements brought about by GenAI are driving the evolution of production methodologies, opening up new opportunities for global economic growth. **Much like conventional AI**, GenAI demonstrates the advantages of **automation**, which can enhance economic productivity. On the one hand, increased automation enabled by AI substitutes

for a portion of human labor in the production process and enhance economic production efficiency⁶. On the other hand, AI imitates and innovates in cutting-edge technologies through automation, which accelerates the development and diffusion of these technologies⁷. **In contrast to conventional AI**, GenAI shows potential for **universality**, indicating collaborations for AI in various areas that will enable it to function in different domains of social and economic activities⁸. GenAI will optimize and reform production processes, management methods, marketing strategies, and other aspects of the market. This will further promote the upgrading of traditional production methods.

GenAI has far-reaching implications for the industrial structure and has distinct sequential effects on various types of industries. With the deep integration of GenAI in various industries, the number of industries that it empowers and reconfigures will continue to increase. Roland Berger's assessment and analysis revealed that

⁴ McKinsey, 'The economic potential of generative AI: The next productivity frontier' (14 June 2023), <<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>> accessed 17 September 2023.

⁵ Goldman Sachs Research, 'Generative AI could raise global GDP by 7%' (05 April 2023), <<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>> accessed 17 September 2023.

⁶ Agrawal, Ajay K., et al. Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth (April 2018). NBER Working Paper No. w24541, p.26.

⁷ WEBB M, The Impact of Artificial Intelligence on the Labor Market (Rochester 2019).

⁸ Partha Pratim RAY, ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope (April 2023), Internet of Things and Cyber-Physical Systems vol.3, p.121-p.154.

GenAI would, at first, have a great impact on **knowledge-intensive industries** such as the Internet, high-tech, financial, and professional service industries by decreasing their costs by 6.5%, 6.8%, and 11.3% respectively. GenAI would then empower **service industries** such as education, communications, medical services, public services, retail, entertainment, and media. GenAI would later impact **traditional industries** such as agriculture, materials, construction, and energy, which are currently not yet highly digitized. Overall, GenAI needs a solid foundation of information and digitalization to realize its potential. And, GenAI is expected to create significant value in R&D design, production and manufacturing, and operations management⁹.

GenAI bolsters the digital economy and introduces a new set of development opportunities across various industries and sectors. The increasing reach of GenAI has facilitated the digital transformation and growth of various industries and sectors. **In the financial industry**, GenAI can aid in creating financial risk profiles and assist in combating money laundering and other financial crimes¹⁰. **In the automobile industry**, GenAI can enhance the efficiency of intelligent voice interaction in vehicles and provide high-quality synthetic data for training automatic driving models. This, in turn, help address data and test challenges associated with the development of automatic driving systems¹¹. Furthermore, multi-modal models are expected to accelerate the application from multi-modal perception to decision planning in end-to-end autonomous driving. **In the media industry**, GenAI can generate text, images, audios, videos,

and other engaging visual content for advertisements based on the text prompts. **In the manufacturing industry**, GenAI can be applied to machine vision, digital twins, autonomous navigation, and other systems to achieve unmanned and intelligent production lines, warehousing, and logistics. **In the agricultural industry**, GenAI can be used to measure crop growth¹², monitor crop diseases, and predict crop yields through large-scale remote sensing models. Overall, the progressive value of GenAI will continue to stimulate quality, efficiency, and dynamics transformation in various industries, promoting high-quality economic development.

(2) Promote Social Progress

GenAI can enhance the effectiveness of city operations services. GenAI can identify problems and optimize strategies through continuous data analysis¹³. For example, it can analyze energy grid data to generate better strategies for optimal energy use. It can also be used in warning systems and to provide comprehensive risk management solutions. Before disasters occur, it can identify the potential patterns of different types of disasters, and provide early warning against disasters through real-time data monitoring and analysis. After disasters occur, it can assist in post-disaster assessment, resource allocation, and rescue planning by analyzing relevant data¹⁴.

GenAI has a positive impact on education. From the perspective of providing opportunities, the benefits of GenAI in information retrieval, learning planning and collaborative creation,

⁹ Roland Berger, "The Dawn of General Artificial Intelligence: The Industrial Impact of Generative Artificial Intelligence Technology", August 2023, p. 3.

¹⁰ Wharton school of University of Pennsylvania, "Does Generative AI Solve the Financial Literacy Problem?" (27 June 2023), <<https://knowledge.wharton.upenn.edu/article/does-generative-ai-solve-the-financial-literacy-problem/>> accessed 18 September 2023.

¹¹ China Enterprise Network, "Application of Generative AI Represented by Chat GPT in the Field of Autonomous Driving" (April 2023), <<http://www.zqcn.com.cn/757/27327.html>> accessed 27 September 2023.

¹² China Science and Technology Information, "Let AI Sink into the Field and Open a New Chapter in Remote Sensing Applications with Large Models" (May 2023), <https://www.thepaper.cn/news-Detail_forward_23222846> accessed 27 September 2023.

¹³ CITNEWS, "How SmartCityGPT Generative Artificial Intelligence Helps the Development of Smart Cities" (August 2023), <<http://www.citnews.com.cn/news/202308/164399.html>> accessed 27 September 2023.

¹⁴ Deloitte, "The ChatGPT era of data and reality integration, exploring new models of smart cities" (July 2023), <<https://www2.deloitte.com/cn/zh/pages/technology-media-and-telecommunications/articles/2023-mwc-chatgpt.html>> accessed 27 September 2023.

among others¹⁵ can help improve the quality of education and teaching¹⁶. From the perspective of promoting equity in education, GenAI can be used to build open educational platforms, provide equal resources for students in different regions, and help bridge the educational gap.

GenAI can promote the expansion and quality improvement of employment. The evolution of GenAI can adjust the employment structure which will create new career choices to include roles such as language model trainers, fact-checkers, robot testing engineers, and data annotators. GenAI can also be used to improve work efficiency by generating documents, processing data, displaying diagrams, and supporting teamwork. The International Labour Organization reported that nearly 427 million jobs are likely to be enhanced by GenAI technology worldwide.

GenAI can drive the development of smart healthcare and geriatric care. In smart healthcare, GenAI applications spread across clinical consultation, medical imaging, clinical diagnosis, and personalized medical intervention scenarios, helping to improve the efficiency and accuracy of many clinical and management applications, improving healthcare services and patient experiences. In smart geriatric care, GenAI can help caregivers adjust their care plans in a timely manner to meet the increasing demand for geriatric care. This can be achieved through the implementation of real-time monitoring systems that analyze the physiological markers of individuals, enabling the timely identification of irregularities.

(3) Support Public Welfare

GenAI helps to create accessible digital envi-

ronments. For example, by utilizing the speech synthesis feature of deep neural networks and voice customization technologies, GenAI can generate voices that are more realistic and offer a wider range of audio books. The cross-modal capabilities of GenAI could be even more beneficial for disabled individuals. Applications such as subtitle glasses and sign language digital humans can help create accessible work environments.

GenAI supports the protection and promotion of global cultures. Relying on 3D scanning and digital storage of artifacts and artworks, GenAI can further apply image generation technologies to reproduce their contents and enhance immersive cultural experiences. It can also apply speech recognition, language synthesis, and virtual human technologies to generate dialect anchors, aiding in dialect preservation and facilitating more dynamic communication.

GenAI boosts global environmental governance and promotes sustainable development. By comprehensively analyzing pollution levels, land surface changes, and climate change data, GenAI can provide suggestions on monitoring, restricting, and implementing measures to mitigate the environmental impacts resulting from human activities. This contribution can help address climate change and promote the sustainability of forests, oceans and wildlife habitats.

(4) Contribute to Scientific Research

GenAI continuously enhances research efficiency. GenAI begins to play an increasingly important role in scientific research, significantly accelerating scientific progress through human-machine collaboration. GenAI can also free researchers from non-essen-

¹⁵ Orly Lobel, THE EQUALITY MACHINE: Harnessing Digital Technology for a Brighter, More Inclusive Future (Public Affairs 2019).

¹⁶ International Telecommunication Union, 'Measuring Digital Development: Facts and Figures 2021', <<https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf>> accessed 15 September 2023.

tial research tasks by assisting in completing fundamental work in academic research, such as collecting materials, analyzing and demonstrating data, and writing paper. GenAI-based academic assistants, for example, provide features such as conclusion extraction, method summary, viewpoint summary, literature review, and abstract integration. These features assist researchers in reading academic literature in a structured and systematic manner, allowing them to quickly capture the research basis, identify the topic of scientific research, and design research proposals¹⁷.

GenAI accelerates scientific discovery and facilitates in-depth exploration in various research directions. GenAI based on deep learning technology, has the potential to effectively leverage the advantageous features of high-dimensional approximation in neural networks. This capability holds promise for addressing the challenges associated with the “Curse of dimensionality” phenomenon, thereby facilitating progress in scientific problem-solving and knowledge discovery endeavors¹⁸. **For molecular structure research,** GenAI can accelerate chemical synthesis throughout the entire process, from molecular and reaction design to condition generation and reaction testing. It can also aid in the design of potential functional molecules, such as pharmaceutical molecule, and their synthesis schemes¹⁹. **For kinetic research,** GenAI can be used in the calculation and simulation of fluid particle models to deduce internal fluid dynamics from external visual observation. Additionally, it can be utilized to determine fluid properties such as viscosity and density through inversion. Therefore, it overcomes the computational limitations of current fluid numerical simulation methods in several

areas, including problem size, simulation speed, model applicability, and accuracy in solving inverse problems. **For the R&D of biomedicines,** GenAI is capable of interacting across different models, covering various scenarios in the pharmaceutical R&D processes. For example, experts can obtain basic information about a specific indication or the progress of pharmaceutical R&D through natural dialogues. GenAI supports multi-modal biomedical data, such as obtaining the complete description information of the molecule by inputting its formula²⁰.



¹⁷ NKI, 'Introduction to CNKI AI Academic Research Assistant' (August 2023), <<http://yuanjian.cnki.com>> accessed 18 September 2023./

¹⁸ The term “Curse of dimensionality” describes the rise in computing’s dimensional complexity or number of variables. See Weinan E’s speech at 2021 World Internet Conference, <https://www.thepaper.cn/newsDetail_forward_13498964> accessed 12 October 2023.

¹⁹ People’s Daily, ‘Shanghai Jiaotong University Open Source Launches Magnolia Science Big Model’ (July 2023), <<https://news.sjtu.edu.cn/mtjj/20230710/185884.html>> accessed 16 September 2023.

²⁰ China Net Science, ‘Shuimu Fenzi Release a Hundred Billion Parameter Big Model of the Biomedical Industry, Launch Pharmaceutical Research Assistant ChatDD’ (September 2023), <http://science.china.com.cn/2023-09/22/content_42532416.htm> accessed 19 September 2023.



04/ Challenges Brought by GenAI

(1) Inherent Technical Risks Cause Security and Safety Concerns

GenAI technologies, when iterated and upgraded, entail increased technical security and safety risks. In terms of data, data feeding is accompanied by problems such as bias in values, disclosure of privacy, and pollution of data. **Firstly**, the inherent bias of the training data leads to bias in the model. Many experiments conducted by global scientific research institutions has shown that large models, even manually annotated, contain biases during application, such as gender and racial discrimination. According to the research report released by Microsoft on GPT-4, when generating profession-

al gender descriptions, large models will further expand the existing bias in the data set, leading to significant gender bias²¹. **Secondly**, the large amount of training data increases the risks of data security and privacy protection. The training of large models relies on huge volumes of data, which brings many challenges to the compliance review of data sources. Furthermore, large models also pose the risk of data leak by inappropriate user input. Samsung reported three data leak incidents by its employees within 20 days after it began to use ChatGPT, and the disclosing contents included sensitive information such as measurements of semiconductor equipment, yields/defects, and internal meetings messages²². **Thirdly**, AI-generated data will pollute the training data. “Model collapse” has been found to occur if AI-generated data is used as a corpus to train a GenAI model. A Cambridge scholar states that GenAI is destroying the internet environment while bringing convenience²³.

In terms of algorithms, the generative charac-

²¹ Microsoft Research, ‘Sparks of Artificial General Intelligence: Early experiments with GPT-4’ (22 March 2023), <<https://arxiv.org/pdf/2303.12712v1.pdf>> accessed 16 October 2023..

²² Ripple Effect, ‘Save ‘Lost AI’? ChatGPT’s Gender Bias and ‘Feeding’ Ethics’ (February 2023), <https://www.thepaper.cn/newsDetail_forward_22047029> accessed 18 September 2023.

²³ Metaverse Post, ‘Ross Anderson Discusses AI Model Collapse as a Growing Problem in Online Content’ (14 June 2023), <<https://mpost.io/ross-anderson-discusses-ai-model-collapse-as-a-growing-problem-in-online-content/>> accessed 15 September 2023.

teristics and security vulnerabilities of algorithmic models bring risks such as “hallucination” or false information and cyber-attacks. **Firstly**, the imitative characteristics of generative models can generate “hallucination” or false information. Since GenAI is based on imitating the training data rather than understanding it, it has the potential to produce false, inaccurate, and unreal information, resulting in what can be referred to as “hallucination”. NewsGuard tested ChatGPT by prompting it with 100 samples from the Misinformation Fingerprints of false narratives. The result was that in 80 out of the 100 prompts, ChatGPT delivered false and misleading claims²⁴. **Secondly**, the security vulnerabilities of algorithmic models trigger the risk of cyber-attacks. Through carefully manipulated input, attackers could exploit or gain control of GenAI back-end systems. There has already been an instance of hackers stealing ChatGPT accounts by modifying the configuration of the Web test suite, SilverBullet²⁵. Although OpenAI has launched a “Bug Bounty Program” that offers up to USD 20,000 to users reporting vulnerabilities, GenAI is still likely to face serious cyber-attack risks over time.

In addition, the “root attribute” of the underlying GenAI models induces route-dependency risk. At present, the GenAI industrial ecology is starting to develop, and the significant costs and substantial investments associated with the underlying large models create significant technical barriers and provide strong competitive advantages that solidify their “fundamental role” as infrastructure. Midstream and downstream product applications developed and deployed based on a large model may spread the inherent risks along its links. In February 2023, the Brookings Institution reported that downstream devel-

opers, who are excluded from the development of original GenAI models, may adapt and integrate these models into other software systems. This could potentially increase the risk of software errors and loss of control, as these developers may not have a full understanding of the entire systems²⁶.

(2) Shifts in the human-machine relationship heighten ethical lapses in use of technology

The reshaping of the human-machine relationship by GenAI may bring ethical misconduct amplified by GenAI. People could become intellectually dependent on GenAI due to its powerful task processing ability. Excessive dependence on the answers provided by GenAI could deprive people of the opportunity to develop their own perception and logical abilities, such as observation, and understanding, induction, and deduction, comparison, and reasoning. People may become too lazy to think and innovate. For example, some students began to rely on GenAI to complete their homework. A professor at Northern Michigan University in the United States discovered an exceptional course paper in his class, only to later realize that it was generated by ChatGPT²⁷. **Potential value deviation or moral defect of GenAI could lead to negative behaviors induced by machines.** In the process of human interaction with GenAI, there have been situations where chatbots have exhibited “emotional” or “aggressive” behavior, and may even have induced human suicide, in part to the lack of ethical considerations in GenAI, which could lead to the generation of harmful stimuli that contradict human ethical values. **GenAI’s strong personified characteristics further impact human subjectivity.** In terms of interpersonal

²⁴ NewsGuard, ‘Despite OpenAI’s Promises, the Company’s New AI Tool Produces Misinformation More Frequently, and More Persuasively, than its Predecessor’ (March 2023), <<https://www.news-guardtech.com/misinformation-monitor/march-2023/>> accessed 15 September 2023.

²⁵ CSO, ‘Stolen ChatGPT premium accounts up for sale on the dark web’ (14 April 2023), <<https://www.csoonline.com/article/575057/stolen-chatgpt-premium-accounts-up-for-sale-on-the-dark-web.html>> accessed 15 September 2023.

²⁶ Brookings, ‘Early thoughts on regulation generative ChatGPT’ (21 February 2023), <<https://www.brookings.edu/blog/techtank/2023/02/21/early-thoughts-on-regulating-generative-ai-like-chatgpt/>> accessed 15 September 2023.

²⁷ The New York Times, ‘Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach’ (16 January 2023), <<https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html>> accessed 15 September 2023.

relationships, frequent interactions with machine could lead to emotional dependence on them and alienation from human social interactions. In terms of human-machine relationships, GenAI could progressively replace human functions, impacting existing divisions of labor and perhaps subjecting people to the “sway” of machines.

(3) Technological Transitions Bring Challenges to the Development of Human Society

GenAI highlights uneven development and widens development gaps. For multi-lingual situations, the available text to train models is very limited. For example, although the Arabic language is spoken by over 420 million people, Arabic language resources and tools in software applications fall short of what is needed, as most software applications primarily support English²⁸. This limited access to language data becomes a significant obstacle for countries or populations that use minority languages to develop GenAI. **GenAI development is likely to widen educational inequities and deepen the digital divide in education.** Since the application of GenAI raises high requirements on hardware facilities and digital literacy, disparities in economic conditions between regions and groups may deepen gaps in educational resources distribution, leading to unequal learning opportunities²⁹. **GenAI impacts the structures of labor and employment and exacerbates social inequality.** Goldman Sachs estimated that while creating new jobs, GenAI could expose the equivalent of 300 million full-time jobs to automation³⁰, with potentially the greatest impact on physically intensive occupations. As a result, more “useless classes” will emerge, and social stratification will become more apparent.

GenAI may impact the ecological environment. The computational and environmental costs of model training are proportional to the size of the model. If plenty of energy is consumed during repeated training, it not only results to resource wastage, but also contributes to an increase in carbon emissions. According to a scholar affiliated with Boston University, it has been observed that the GPT-3 model, which consists of 175 billion parameters, consumes an amount of energy comparable to 1287 megawatt hours of electricity. Additionally, this model is responsible for generating of approximately 552 tons of carbon dioxide emissions³¹. **GenAI is impacting the intellectual property system.** Whether GenAI is qualified for authorship or can obtain copyright for its content are still open questions. Moreover, any unauthorized use of articles or codes, or any violation of open-source licenses, could constitute an infringement on intellectual property rights.



²⁸ United ARAB Emirates, '100 Practical Applications and Use Cases of Generative AI' (April 2023), <https://ai.gov.ae/wp-content/uploads/2023/04/406.-Generative-AI-Guide_ver1-EN.pdf> accessed 15 September 2023.

²⁹ Walczak Krzysztof, Cellary Wojciech, Challenges for higher education in the era of widespread access to Generative AI (April 2023), Economics and Business Review vol.9(2), p.71-p.100.

³⁰ Goldman Sachs Research, 'The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)' (26 March 2023), <https://www.key4biz.it/wp-content/uploads/2023/03/-Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf> accessed 17 September 2023.

³¹ Kate Saenko, 'Is generative AI bad for the environment? A computer Scientist explains the carbon footprint of ChatGPT and its cousins' (23 May 2023), <<https://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096>> accessed 11 October 2023.



05 /

Global Efforts to Develop Responsible GenAI

(1) Efforts Made by International Organizations

International organizations actively explore principles and norms for AI development. In May 2019, the **Organization for Economic Co-operation and Development (OECD)** published the first set of intergovernmental principles for AI. In June 2019, the **G20** further elaborated on the OECD AI Principles to develop the “G20 AI Principles”, to foster public trust and confidence in AI technologies and fully realize their potential. The “G20 AI Principles” include “inclusive growth, sustainable development and well-being”, “human-centered values and fairness”, “transparency and

explainability”, “robustness, security and safety” and “accountability”, which have been universally recognized in the international community. In June 2021, the **World Health Organization (WHO)** released *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, which presents six tenets and shares a set of recommendations to ensure that AI respects the public interest of all countries. The six tenets include protecting human autonomy, promoting human well-being and safety and the public interest, ensuring transparency, explainability and intelligibility, fostering responsibility and accountability, ensuring inclusiveness and equity, and promoting AI that is responsive and sustainable. In November 2021, the **United Nations Educational, Scientific and Cultural Organization (UNESCO)** adopted the first global AI ethical instrument, *Recommendation on the Ethics of Artificial Intelligence*, which proposes four values to be respected in the development and application of AI: “respect, protection and promotion of human rights and fundamental freedoms and human dignity; environment and

ecosystem flourishing; ensuring diversity and inclusiveness; living in peaceful, just, and interconnected societies”. In October 2023, the **United Nations** (UN) Secretary-General convened a High-level Advisory Body on AI to undertake analysis on addressing AI risks and opportunities, and advance recommendations for the international governance of AI.

International organizations are highly concerned about developing responsible GenAI. In April 2023, the OECD’s *AI Language Models* proposed policy considerations related to GenAI such as language models, building upon the existing governance principles benchmark. In May 2023, the UN released a policy brief on *A Global Digital Compact*, outlining four objectives to realize agile governance of AI and other emerging technologies, including ensuring that the design and use of these technologies are transparent, reliable, safe and under accountable human control; make transparency, fairness and accountability in AI governance; integrating international guidance, national regulatory frameworks, and technical standards to establish a framework for agile governance of AI; in the case of regulators, coordinate across multiple aspects to ensure the alignment of emerging digital technologies with human values. In June 2023, the **World Economic Forum** (WEF) released the *Presidio Recommendations on Responsible Generative AI*. This report presents a set of 30 action-oriented recommendations for various stakeholders. The recommendations address three key themes: responsible development and release of GenAI; open innovation and international collaboration; and social progress. The objective is to foster accountable and inclusive processes for GenAI development and deployment, thereby increasing trust and transparency as GenAI systems continue to

spread. In June 2023, the **World Internet Conference** launched the “AI for Social Good Action Plan (2023-2025)”, calling for the utilization of AI technologies to enhance human welfare. In September 2023, UNESCO released *the Guidance for Generative AI in Education and Research*. Based on a human-centered principle, this document lists seven key steps for countries to regulate the application of GenAI in education, including the development or adaptation of existing policies and regulations on data protection, AI policies, AI ethics, copyright law, and GenAI regulation, etc., develop guidelines for the use of GenAI in education and research, and review its long-term influence in this field³².

(2) Efforts Made by Major Countries and Regions

For AI ethics, major countries and regions are exploring the law-based governance and regulation while also encouraging innovation. *The Ethics Guidelines for Trustworthy AI*, released by the **European Union (EU)** in 2019, prescribes seven ethical principles. These principles include human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability³³. In the upcoming *Artificial Intelligence Act*, the EU intends to adopt a risk-based classification regulation approach to promote responsible research and innovation. In 2019, **China** released the *Governance Principles for New Generation AI - Developing Responsible AI*. These principles emphasize eight principles including harmony and friendliness, fairness and equity, and inclusiveness and sharing, among others. The aim is to ensure the development of

³² UNESCO, ‘Governments must quickly regulate Generative AI in schools’ (September 2023), <<https://www.unesco.org/zh/articles/jiaokewenzuzhigeguoxu-jinkuaguidanfashengchengshirengongzhinengdexiaoyuanyingyong>> accessed 17 September 2023.

³³ European Commission, ‘Ethics guidelines for trustworthy AI’ (8 April 2019), <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>> accessed 17 September 2023.

safe, reliable, and controllable AI. To meticulously implement these principles, China released the *Ethical Norms on New Generation AI* in 2021. These norms put forward 18 specific ethical requirements for the management, R&D, supply, and use of AI. China declared that it would “stand ready to increase exchanges and dialogue with other countries and jointly promote the sound, orderly and secure AI development in the world”³⁴. At the Third Belt and Road Forum for International Cooperation in October 2023, China also unveiled the *Global AI Governance Initiative* at the forum. The *Artificial Intelligence and Data Act* proposed by **Canada** in June 2022 intends to build a responsible AI framework that guides AI innovation in a positive direction. It requires high-impact AI systems to follow principles such as human oversight and monitoring, transparency, fairness and equity, safety, accountability, validity, and robustness³⁵. The *Blueprint for an AI Bill of Rights*, released in October 2022 by the Office of Science and Technology Policy of the White House, the **United States(U.S.)**, is accompanied by five principles of safe and effective systems, algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives, consideration, and fallback, as well as practice guidelines³⁶; the U.S. National Institute of Standards and Technology issued the *Artificial Intelligence Risk Management Framework* and the supporting playbook in January 2023, which describes the core functions of the Framework, namely, govern, map, measure, and manage, to help organizations address the risks of AI systems in practice³⁷. In October 2023, the

U.S. president issues *Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, to establish new standards for AI safety and security³⁸. In March 2023, the **United Kingdom(U.K.)** Department for Science, Innovation and Technology released the white paper called *A Pro-innovation Approach to AI Regulation*, which includes five principles: safety, security and robustness, appropriate transparency and explainability, fairness, accountability and governance, and contestability and redress to guide and inform the responsible development and use of AI³⁹. In November 2023, the U.K. hosted the AI Safety Summit, during which 28 countries including China, the U.S., and the U.K., as well as the EU, jointly signed the “The Bletchley Declaration”⁴⁰.

For GenAI ethics, major countries and regions have shown swift responses and exercised cautious guidance. The rapid development of ChatGPT and other GenAI technologies has a significant impact on the way in which AI systems are constructed and deployed and stirs deeper concerns. In the revised draft of *Artificial Intelligence Act*, the **EU** imposes a transparency requirement on GenAI providers, requiring disclosure of whether the content is generated by AI. It also requires methods in distinguishing between deepfake images and real images, and the disclosure of the copyright of the training data⁴¹. In both July and September 2023, the U.S. Government convened prominent AI corporations. At the convening, these corporations voluntarily promised to foster the secure, dependable, and reputable advancement of AI

³⁴ Xi Jinping’s Keynote Speech at the 3rd Belt and Road Forum for International Cooperation.

³⁵ Government of Canada, ‘The Artificial Intelligence and Data Act (AIDA) – Companion document’ (13 March 2023), <<https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>> accessed 17 September 2023.

³⁶ White House, ‘Blueprint for an AI Bill of Rights’ (October 2022), <<https://www.whitehouse.gov/ostp/ai-bill-of-rights/>> accessed 17 September 2023.

³⁷ NIST, ‘Artificial Intelligence Risk Management Framework’ (January 2023), <<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>> accessed 17 September 2023.

³⁸ White House, ‘FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence’ (30 October 2023), <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>> accessed 6 November 2023.

³⁹ Gov. UK, ‘A pro-innovation approach to AI regulation’ (3 August 2023), <<https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>> accessed 17 September 2023.

⁴⁰ Gov. UK, ‘The Bletchley Declaration by Countries Attending the AI Safety Summit’ (1 November 2023), <<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>> accessed 6 November 2023.

⁴¹ European Commission, ‘LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS’ (April 2021), Article 52.

⁴² White House, ‘FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI’ (12 September 2023), <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>> accessed 17 September 2023.

technology⁴². Additionally, the government declared a range of measures, which included conducting a public evaluation of current operational AI systems⁴³. The U.S. President's Council of Advisors on Science and Technology has established a GenAI working group to provide guidance on the safe, fair, and responsible development and deployment of GenAI related technologies⁴⁴. In June 2023, the U.K. Cabinet Office noted the need to identify the opportunities and risks of new technologies and set out general principles for the use of GenAI by UK civil servants in *Guidance to Civil Servants on Use of Generative AI*⁴⁵. In addition to the previous *Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services and Provisions on the Administration of Deep Synthesis of Internet-based Information Services*, **China** also specially promulgated the *Interim Measures for the Administration of Generative Artificial Intelligence Services* in August 2023. Upholding the principles of putting equal emphasis on development and security, integrating innovation with rule-of-law governance, it introduces effective measures to encourage the innovative advancement of GenAI. In August 2023, **Canada** identified the elements in a code of practice for GenAI based on the principles set forth in *AIDA*, which provide specific references for developers, deployers and operators⁴⁶. Later, Canada released *Guide on the use of Generative AI*. To maintain public trust and ensure the responsible use of GenAI tools, federal institutions are required to align with the "FASTER" principles: Fair, Accountable, Secure, Transparent, Educated, and Relevant. This document provides preliminary guidance to federal institutions on

their use of GenAI tools⁴⁷. At the same time, the U.K., France, Japan, South Korea and other countries are also actively committed to AI governance.

(3) Efforts Made by the Industry

The global industry is actively practicing responsible AI by establishing codes of conduct and guidelines, creating ethics committees, and developing governance tools. In terms of codes of conduct and guidelines, companies such as IBM, Microsoft, Google, Baidu, Tencent, Alibaba, Ant Group, SenseTime and 360 have introduced their corporate AI principles, covering benefits to society, security, privacy protection, fairness, transparency, explainability, controllability, and accountability. In terms of industrial organization, the Artificial Intelligence Industry Alliance of China released the *Pact on Self-Discipline of Artificial Intelligence Industry*, and the China Academy of Information and Communications Technology released the *White Paper on Trustworthy Artificial Intelligence*. On the basis of soliciting feedback from the industry, the Chinese Academy of Social Sciences Law Institute team has recently published the *Model Law on Artificial Intelligence (Expert Draft Proposal) v.1.1*⁴⁸. **In terms of governance organization**, many large technology companies, including IBM, Microsoft, Google, Lucid AI, SenseTime, Alibaba and MEGVII have established their ethics committees. Among them, Microsoft has three parallel institutions responsible for AI ethics affairs, and IBM's AI Ethics Board supports all project teams in the company to adhere to AI ethics principles. For SenseTime, it has estab-

⁴³ White House, 'FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety' (04 May 2023), <<https://www.whitehouse.gov/ostp/news-updates/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/>> accessed 17 September 2023.

⁴⁴ White House, 'PCAST Working Group on Generative AI Invites Public Input' (13 May 2023), <<https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/>> accessed 17 September 2023.

⁴⁵ Gov.UK, 'Guidance to civil servants on use of generative AI' (29 September 2023), <<https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai>> accessed 17 September 2023.

⁴⁶ Government of Canada, 'Canadian Guardrails for Generative AI – Code of Practice' (16 August 2023), <<https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>> accessed 17 September 2023.

⁴⁷ Government of Canada, 'Exploring the future of responsible AI in government' (6 September 2023), <<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html>> accessed 17 September 2023.

⁴⁸ IOLAW, 'The Chinese and English versions of the MODEL LAW ON ARTIFICIAL INTELLIGENCE (EXPERT DRAFT PROPOSAL) v.1.1 have been released globally' (September 2023), <http://iolaw.csn.cn/zxz-p/202309/t20230907_5683898.shtml>, accessed 20 September 2023.

lished an ethical risk audit team to conduct ethical reviews of all its products. **In terms of implementation tools**, enterprises also actively promote a technology-based and project-based approach to ethical governance principles of science and technology by developing diverse ethical governance tools and internal action guides. IBM, for example, developed several trustworthy AI tools in 2018 to assess and test the fairness, robustness, explainability, accountability, and value consistency of AI products during the development process⁴⁹. Other enterprises such as Microsoft, Google, JD.com, HUAWEI, Baidu, Tencent, Ant Group, MEGVII, SenseTime and Yinxiang Biji are also actively engaged in relevant practices.

Under the wave of GenAI, enterprises actively explore new solutions to deal with new risks.

To enhance security and robustness, Cisco uses security, data breach and privacy incident response systems to manage AI incidents relating to bias and discrimination, and the company reports investigation findings and remedial steps to broader stakeholders⁵⁰. Ant Group launched an automated intelligent security and risk identification system named "Yi Jian 2.0", encompassing GenAI safety, robustness, explainability evaluations for large-scale models. To protect data privacy, GitHub Copilot provides users with the option to use standard configuration files which are stored on local machines to reduce the need to access cloud APIs, thereby protecting user privacy. SenseNova, a large model set developed by SenseTime, includes audit nodes for personal information protection at the stages of design, coding, testing, and delivery, to subject the processing of personal information to necessary checks⁵¹. **To improve transparency**, IBM has launched the AI FactSheets 360 website, which

provides “fact sheet” assembly methods based on the important features of AI models (including purpose, performance, datasets, features, etc.), to help improve AI transparency across the industry⁵². **To strengthen explainability**, OpenAI uses GPT-4 to automatically write explanations for the neural network behavior of GPT-2 and score these explanations, in an attempt to promote the research on explainability and comprehensibility of models. **To guarantee fairness and inclusive**, Google recently introduced a novel approach called the "LASSI" representation learning method. This method aims to assess the individual fairness of high-dimensional data and enhance the processing of user data in an objective manner⁵³. **To ensure value alignment**, OpenAI created the Superalignment team to achieve the goal of aligning AI systems with human values within four years. **To identify generated contents**, the Meta open-source digital watermark “Stable Signature” is a novel technique that enables the direct embedding of digital watermarks into images generated by AI, which improves the security of GenAI. Douyin advocates a unified AI-generated content identification capability across the platform to help creators mark content for differentiation⁵⁴.



⁴⁹ IBM Research, 'Trustworthy AI' (September 2023), <<https://research.ibm.com/topics/trustworthy-ai>> accessed 17 September 2023.

⁵⁰ CISCO, 'The Cisco Responsible AI Framework' (2022), <https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-responsible-artificial-intelligence-framework.pdf?CCID=cc000742&DTID=esootr000875> accessed 17 September 2023.

⁵¹ SenseTime, 'White Paper on Principles and Practice of Large Model Governance', August 2023.

⁵² IBM Research, 'IBM Artificial Intelligence Pillars' (30 August 2023), <<https://www.ibm.com/policy/ibm-artificial-intelligence-pillars/>> accessed 17 September 2023.

⁵³ Bhuvana Kamath, 'Generative AI Is Biased. But Researchers Are Trying to Fix It', <<https://analyticsindiamag.com/generative-ai-is-biased-but-researchers-are-trying-to-fix-it/>> accessed 11 October 2023.

⁵⁴ Douyin, 'Douyin initiates 11 rules to regulate AI-generated content for platforms', 9 May 2023.



06 / Consensus on Developing Responsible GenAI

(1) General

01 Developing responsible GenAI should consistently commit to enhancing human well-being, adhering to a people-centered approach, and promoting the sustainable development of the economy, society, and ecological environment. It is important to accurately recognize the immense potential and possible risks of GenAI. By adhering to a holistic approach that harmonizes development with safety and security, balances innovation with ethical considerations, and fairly evaluates benefits and risks, we can promote the responsible development of GenAI. On the one hand, innovation-driven, sustainable, inclusive, and open development of GenAI should be promoted, and the efficient computing power, high-quality data, innovative algorithm, diverse talents, and an open ecosystem associated with GenAI should be

enhanced. On the other hand, GenAI characterized by reliability and controllability, transparency and explainability, data protection, diversity and inclusiveness, accountability, and value alignment should be developed with a highly responsible attitude.

(2) Promote the Development of GenAI

02 Actively advocate and prudently promote the sustainable development of GenAI. Firstly, ensure economic sustainability. It is vital to ensure that GenAI can enhance productivity and create employment opportunities, improve the efficiency of resource utilization, achieve a circular economy that integrates digital technologies with the real economy, propel scientific and technological innovation, and facilitate the transformation to an economic structure with higher value added. Secondly, ensure social sustainability. It is crucial to ensure the fair and equitable use of GenAI and to promote co-construction and co-governance at the societal level. Thirdly, ensure environmental sustainability. The development of GenAI should promote the sustainable management and use of natural resources, encourage the use of green energy-driven infrastructure,

and enhance green R&D of models and applications, to reduce greenhouse gas emissions and achieve green development.

03 Create an environment that is conducive to healthy and orderly development of GenAI.

Firstly, establish and optimize relevant ethical principles, laws, and regulations. Legal assessment of the adequacy of the current intellectual property laws should be done with concerned stakeholders to ensure proper attribution of rights to AI-generated objects and to appropriately manage and protect them. **Secondly, establish an inclusive, supportive, forward-looking, and predictable policy environment.** GenAI requires an inclusive innovation environment for incubating its cutting-edge applications, an excellent business environment for deploying them at scale, and a robust regulatory environment to drive economic and social development. **Thirdly, strengthen international communication and cooperation.** The development of GenAI requires all global stakeholders to uphold the principle of consultation, contribution, and shared benefits, to engage in transnational, cross-disciplinary, and cross-cultural communication and cooperation with open and collaborative attitudes and actions, to jointly establish an international assessment and standard system with broad consensus, to ensure that all countries can share in the technological benefits of GenAI.

04 Enhance the capabilities of R&D and scaling the application of GenAI.

Firstly, build open, shared, and inclusively computing resources that are beneficial to all. Efforts should be made to promote the rational allocation and efficient use of computing power, reducing the barriers to scientific and technological innovation. This will enable enterprises of different regions and tiers to access the computational resources they need. **Secondly, promote responsible data sharing.** Efforts should be made to encourage the

sharing and flow of high-quality data, increase the supply of public data resources, ensure the safe sharing and compliant use of data, and strengthen the level of data governance in all fields. **Thirdly, optimize the facility system for algorithm innovation.** Forward-looking planning and overall arrangement for the construction of various platforms and open shared service networks should be enacted, secure open-sourcing of algorithms and basic models should be encouraged, cross-industry and cross-domain collaboration should be strengthened, and the integration of industry, academia, and research should be promoted, to form a virtuous ecosystem for algorithmic innovation. **Fourthly, comprehensively strengthen talent capacity building.** For practitioners, a talent exchange platform should be established to promote mutual learning and knowledge sharing, and multi-level, multi-domain education and training programs should be designed and implemented to enhance communication and learning between technology providers and users in different fields. For the public, science popularization, education, and training should be enhanced to provide accurate knowledge, improve the digital literacy, and promote universal access to GenAI. **Fifthly, enhance the empowerment of critical application areas.** Promote the seamless integration of GenAI with digital scenarios across various industries, enabling application innovation and iteration, and driving the adoption of GenAI in key sectors.

(3) Enhance the Capacity for Responsible Governance of GenAI

05 Develop safe and reliable GenAI to ensure controlled operation throughout its lifecycle.

Firstly, improve the safety, robustness, and generative accuracy of GenAI. Enhance the capabilities of generative models to defend against prompted attacks, and other types of attacks, while continuously improving their robustness and anti-interference capabilities.

Explore technologies or solutions for ensuring controlled content generation to ensure that the generated information is as accurate as possible. **Secondly, ensure humans are well-informed and in control.** It is important to ensure that humans are aware of their interactions with GenAI and that GenAI systems are supervised and controlled by humans in a timely manner. **Thirdly, avoid the abuse and misuse of technologies.** Mitigate excessive dependence on GenAI by users and reduce its negative impact on human innovation and subjectivity. Avoid intentionally or unintentionally using of GenAI to harm society and the public interest.

06 Enhance the transparency and explainability of GenAI systems to improve human understanding and trust. **Firstly, improve transparency.** Encourage disclosure of capabilities and limitations of GenAI systems, as well as decision-making processes and technical intentions, based on safety. Establish external supervision and feedback channels, and continuously make improvements. **Secondly, enhance explainability.** Promote research on the explainability of GenAI, explore robust explanatory technical pathways for adaptive scenarios and risk levels, to enhance human trust and improve application acceptance.

07 Strengthen GenAI data governance and data security, respecting and protecting individual privacy. **Firstly, reinforce data governance.** Problems such as illegal collection, abuse, and breach of training data should be avoided, and effective measures should be taken to improve the quality of training data. **Secondly, enhance the protection of personal information and privacy.** When GenAI training data involves personal data, the owner should be informed and consulted for consent to ensure that the generated content does not infringe on individual privacy. **Thirdly, explore technologies for privacy protection.** When developing

GenAI systems, it is important to explore the use of privacy-preserving computing and other privacy protection technologies to mitigate the risk of data leakage and misuse.

08 Ensure the openness, inclusiveness, and fairness of GenAI. **Firstly, ensure diversity and inclusiveness in technology.** GenAI training data and application scenarios should be diverse to avoid bias and discrimination against specific groups or individuals. **Secondly, promote a fair and equitable access to the technology,** reduce the cost and usage barriers of GenAI, enhance its accessibility and usability, advocate for the sharing of benefits that GenAI brings to human society, promote social fairness and equal opportunities, and bridge the digital divide.

09 Clarify the mechanisms for accountability for GenAI and enhance system traceability. **Firstly, clarify the mechanisms for accountability,** including the appropriate right obligations and accountability mechanisms for various categories of scientific design subjects throughout the life cycle of GenAI. This includes design, training, optimization, deployment, application, and so forth. These mechanisms should ensure liability in the event of any damage occurring. **Secondly, establish a traceability system.** AI ethics committees should be encouraged to establish and improve. Ensure GenAI results are traceable. **Thirdly, explore innovation-friendly governance tools like sandbox,** to provide experimental space for GenAI, fostering responsible innovative exploration.

10 Promote GenAI to better comprehend human intentions, follow human directives, and align with human ethics. **Firstly, explore research on value alignment.** Strengthen GenAI value alignment theory exploration, technological research, and R&D of tools, and improve the ability of humans to design, understand and supervise GenAI models. **Secondly, enhance**

value alignment technologies. Improve the quality of GenAI's training data, adopting manual or automated detection, red team testing, watermarking, content filtering, and other methods to improve its consistency with human values.

Attachment: Exploration of Industrial Applications for Developing Responsible Generative Artificial Intelligence

(1) New Paradigms of Technology Applications Resulting from GenAI

Microsoft reshapes software development with an AI-driven and codeless intelligent approach. Power Platform launched by Microsoft popularizes development activities by adding the “intelligent Copilot” to Power Apps, Power Automate and Power Virtual Agents, allowing more people to create innovative solutions by using natural language. For example, users simply need to imagine a solution and describe it in simple and everyday language, and Copilot can implement the solution and provide intuitive and intelligent low-code experience.

Stable Diffusion, an open-source image generation model, crafts high-quality image work. Stability AI, a UK-based company, has completed the development of the initial version of the Stable Diffusion model by joining forces with LudwigMaximilian University of Munich and University Heidelberg in Germany, as well as AI company Runway. Stable Diffusion, positioned alongside proprietary commercial image generation models like Midjourney and the Dall-E, is a top-tier open-source model with a broad reputation in the global generative AI open-source community. With its powerful image generation capabilities and open-source nature, it has found widespread application worldwide.

Baidu has developed Comate, an intelligent coding assistant tool, to improve production efficiency. This tool is designed to improve coding efficiency, reduce errors, simplify the writing process of test cases, and improve the efficiency and reliability of the software development process. For example, it can automatically generate code snippets based on the natural language description or comment provided by the user, which improves the coding efficiency and reduces errors otherwise caused by manual coding. Based on the code snippet selected by the user, it can automatically generate unit test cases, which saves the time that developers spend writing test cases, ensures comprehensive test coverage, and improves the code quality.

IBM creates a full-stack enterprise-level AI solution that provides end-to-end technology and service support for enterprise digitalization and intelligent transformation. The solution provides a secure, reliable, and high-performance arithmetic and storage environment for AI model training, tuning, reasoning, and other tasks by integrating heterogeneous infrastructures, including IBM mainframes, LinuxONE, and software-defined storage. IBM watsonx builds an enterprise-grade AI and data platform: watsonx.ai provides full-lifecycle oriented enterprise AI models and application development capabilities; watsonx.data provides transparent access to enterprise internal and external data in the lakehouse data platform; watsonx.governance helps enterprises build responsible, transparent and explainable AI application and data governance. In addition, IBM has established the Generative AI Center of Excellence to provide enterprises with a variety of business solutions, including digital

labor, sustainability, intelligent operations and maintenance, code empowerment, security compliance and application modernization, to accelerate the transformation of generative AI into the core productivity of enterprises.

360 launched AI Box plug-in empowers a variety of business scenarios and makes AI easy to use. AI Box plug-in is a special resource/service plug-in covering multiple industries and scenarios. AI Box can assist in generating a more professional and richer scenario resource form, so as to provide more choices and better services for enterprises.

Tencent fully applies large language model technology to practical scenarios to help improve digital productivity. Tencent's large language model "Hunyuan" is based on independent research and development and has a parameter scale of over one hundred billion, with powerful Chinese authoring capabilities, logical reasoning capabilities in complex contexts, and reliable task execution capabilities. On meeting, tasks such as information extraction and content analysis can be accomplished with simple commands, improving the efficiency of information flow. In the workplace, intelligent text creating is supported to facilitate collaboration. On advertising, intelligent advertising material creation is supported to help merchants improve service quality and service efficiency.

Huawei Cloud CodeArts Snap is an intelligent assistant for developers that is powered by the Pangu R&D model. CodeArts Snap automatically generates code, unit test cases, and test scripts. Complete code logic can be generated based on natural language, significantly improving software development efficiency and reducing code errors and vulnerabilities. CodeArts Snap can also intelligently interpret, optimize, debug, review, modify, and add comments to

code based on the intelligent Q&A function, helping developers quickly solve technical problems and improve code readability. In addition, CodeArts Snap can identify requirements, commit code, and run a pipeline, improving collaboration efficiency during software development.

Zhipu AI released an electronic signing generative AI product in the SaaS industry. Based on its self-developed GLM-130B large language model with hundreds of billions of parameters, Zhipu AI has developed the intelligent contracting product "Hubble". The product of Hubble can, under the authorization of the user, retrieve the electronically signed contract through continuous dialogue, and perform functions such as profile interpretation, key content identification, filtering and identification, classification and summarization, helping the user to reduce cost and increase efficiency.

The Technology Innovation Institute (TII) in Abu Dhabi, open-sources an over 100 billion parameter scale of large model with user-friendly examples. In September 2023, TII launched the world's first open-source large model above 100 billion parameter size, Falcon 180B. Researchers also released a chatbot model, Falcon-180B-Chat. This model, fine-tuned on dialogue and command datasets and blended with several large-scale conversation datasets, enhances its comprehension of user text prompts, yielding more fluent textual content.

(2) Business Development Empowered by GenAI + Finance

Citibank adopts IBM's enterprise AI solution to realize the digital and intelligent transformation of its financial business. Citibank completed 5,000 compliance checks by using an advanced analytics solution developed by IBM, saving 30,000 hours that 2,500 auditors would

have otherwise spent on internal audits. At the same time, IBM has also created an AI innovation space for Citibank to promote continuous innovation in the application of AI on the new audit platform. By introducing watsonx, Citibank continues exploring AI-enabled internal control to further realize the intelligent transformation of audit work.

Bank of Communications and Tencent jointly build an intelligent platform. This AI intelligent platform is a unified image recognition platform that integrates new technologies such as computer vision, machine learning and AI to provide the client with enterprise-level one-stop services. It supports customization of application scenarios and can be deployed on mobile, PC, and server terminals to help quickly release efficient and available recognition services.

Lloyds Bank uses GenAI to improve customer services. Lloyds Bank uses foundation and GenAI models across the enterprise to capture ever-changing customer needs, enhance customer interactions, and reduce the manual efforts required to manage, train, and execute AI-driven interaction processes. At the same time, empowered by GenAI, Lloyds Bank is able to acquire important data and improve productivity.

Ant Group launches an intelligent financial assistant “Zhixiaobao 2.0” to provide reliable financial services. Zhixiaobao 2.0 is a new-generation intelligent financial assistant developed by Ant Group based on its self-developed financial model to provide users with financial services and professional investment advice, in addition to companionship and communication. Zhixiaobao 2.0 emphasizes the suitability and security of financial products, helps address the information gap, improves user experience, and boosts continuous innovation and progress in the financial field.

SenseTime uses GenAI products to empower financial business development. SenseTime provides cross-verification of crop underwriting data based on an AI-based remote sensing large model for insurance companies in China. This improves the risk assessment and claim settlement service systems of the insurance company. Compared with the traditional manual identification solution, this model improves the efficiency and serves as a demonstration of technological innovation and application in the agricultural insurance sector.

Tencent utilizes generative AI capabilities to empower the development of finance and other businesses. By using Tencent Cloud's industry-specific large language model solutions, enterprises can easily access one-stop services such as pre-training, fine-tuning, and application development. On the TI platform, built-in high-quality large language models can be combined with an enterprise's own scenario data to generate exclusive models. Customers can customize different parameters and model specifications on-demand, improving the efficiency of their services.

Huawei Cloud Pangu Finance Model equips every employee in financial service organizations with an intelligent assistant. Huawei Cloud worked with the Software Development Center of Industrial and Commercial Bank of China (ICBC) to develop branch office and customer service assistants. By leveraging the generation capability of foundation models and search technology, these intelligent assistants automatically generate service workflows and guides for the counter staff and provide the retrieval source to ensure the reliability of the generated content, simplifying service request handling and improving customer experience. Pangu Finance Model covers risk control, marketing, investment research, movable property

pledges, claim settlement, and customer service. It has comprehensive perception, cognition, decision-making, prediction, and generation capabilities, giving every employee an AI assistant and promoting inclusive finance.

(3) Inclusive Science and Education Supported by GenAI + Education

iFlytek empowers education with GenAI. By holding the tour exhibition of “Science and Technology Caravan” across the country, iFLYTEK brings the most cutting-edge scientific installations, the most interesting interactive experiences and authoritative scientific knowledge about AI and other technologies to teenagers in many underdeveloped areas. This initiative stimulates teenager’s interest in science, develops their scientific spirit, and promotes the development of popular science education. The iFLYTEK Spark APP provides oral English dialogues in diverse settings, multilingual translation, pronunciation evaluation, grammar correction, word searching and other services. It also provides AI-enabled oral teachers for college students and business professionals. The AI stress relieving planet in iFlytek's mental health education solution pioneered the human-machine dialogue stress relieving mode, which can provide high-quality psychological counseling to students at any time.

Intel nurtures next-generation technologists by driving AI accessibility. Intel has expanded its Digital Readiness Programs globally by collaborating with 27 national governments, enabling 23,000 institutions and training more than 5.60 million people around the world. By AI education and building digital readiness, Intel prepares all the students for digital transformation and drives inclusive, accessible, and responsible AI innovation.

SenseTime invents a new AI-enabled education model to promote the development of

inclusive education. SenseTime provides an age-appropriate AI education platform (SenseStudy) for colleges and universities that covers all aspects of the industry chain of education and realizes closed-loop integration. By providing high-quality AI-enabled educational courses, products and services for schools in all phases of studying, this platform will promote the popularization and inclusive development of AI in the field of education. For vocational education, SenseStudy and Shenzhen Institute of Information Technology jointly founded “Shenzhen Institute of Information Technology - SenseTime AI Industry College”, which adopts the “AI + Mobile Internet” and “AI + Augmented Reality” interdisciplinary professional talent training models. By building comprehensive partnership in college-wide general education, AI discipline construction, development of teaching materials, internship in enterprise production, and social training, the partners aim to develop new AI-enabled interdisciplinary talent training models.

(4) Digital Transformation Accelerated by GenAI + Industry

Mitsui chemicals explores new GenAI applications to boost digital transformation. Mitsui and IBM Japan, Ltd. cooperate in exploring new applications of generative pre-training Transformer (GPT) to boost digital transformation of business and further increase the sales and market shares of Mitsui’s products.

SenseTime supports the industrial application of GenAI in the power sector, helping enterprises with cost reduction and efficiency improvement. SenseTime cooperates with leading power companies in China to provide large model capabilities and services in three aspects: (1) Applying “SenseChat” in customer service scenarios to reduce costs and increase efficiency; (2) Relying on multimodal large models to identify and judge long-tail faults and responsible

defects in the open world; (3) Jointly studying the application of an intelligent decision-making system to the intelligent dispatching of the power system.

Microsoft improves industrial efficiency through GenAI collaborative capability. Siemens and Microsoft cooperate in improving industrial productivity with GenAI. The collaborative capability of GenAI is used to assist industrial enterprises to continuously improve efficiency and drive innovation throughout the life cycles of products.

iFLYTEK utilizes GenAI to drive the intelligent upgrading of the manufacturing industry. Based on the core technology foundation of the Spark Cognition Large Language Model, combined with industrial scenarios, iFLYTEK create the Antelope Industrial Big Model, provide professional suggestions for enterprise needs, intelligently match solutions, service providers, experts and other resources, and create a “sustainable enterprise brain” throughout the entire process of research and development, production, sales, and service, helping enterprises efficiently accumulate scenario knowledge and improve operational efficiency.

360 joint industry chain leading enterprises jointly initiated the establishment of GPT industry alliance. It will be through the development of 100 GPT vertical industry partners, hand in hand to develop 1,000 GPT application eco-partners, for millions of customers to provide GPT one-stop service, docking in the upstream and downstream of the GPT industry to cover. At the same time, GPT Industry Alliance will also open API and plug-in system for its members, empowering the intelligent upgrading of applications in different industry segments.

Huawei Cloud Pangu Manufacturing Model offers a fantastic production scheduling solu-

tion for manufacturing and supply chain planning. Pangu Manufacturing Model was pre-trained on data about parts and components, business processes, and rules from Huawei's production line, allowing it to quickly identify a globally optimized solution for improved efficiency and reduced costs through complex simulation and computation. This is a huge improvement over conventional methods for production planning. This model ignited revolutionary changes in manufacturing by providing five major AI skills for manufacturing, including manufacturing requirement analysis, auto generation of process standards documents, auto generation of industrial software code, visual recognition for the production and supply phases, and intelligent planning for the production and supply phases. These skills cover core phases of R&D, production, and supply.

(5) Transportation Convenience Improved by GenAI + Transportation

Intel unleashes the potential of GenAI technologies to enhance 3D experience and promote Automatic Driving technology development. Intel launched CARLA, an open-source urban driving simulator that supports the development, training, and validation of automatic driving systems. By using GenAI, it can improve the realism and naturalness of the scenes around the driver and promote the development and innovation of automatic driving technologies.

PCITECH explores innovative applications of GenAI in the field of transportation, making travel and life more convenient. With respect to the subway scenario, a digital virtual employee of the subway is built to empower the customer service center of the subway station. Leveraging the key capabilities of large language models, such as semantic understanding, context association, answer generation and summary, and self-learning, etc., the virtual employee could

communicate with passengers, answer their inquiries, and guide them to handle tickets.

(6) Inclusive Medical Care Contributed by GenAI + Medical Care

Alibaba DAMO Academy uses GenAI to assist in diagnosis of different corneal diseases. The research team took numerous images of corneal diseases with the help of slit lamps and develop a unique corneal disease diagnosis method. On this basis, they developed the “deep feature learning and recognition algorithm for corneal disease sequence”. The existing special ophthalmologic examination equipment became intelligent after they were embedded with the intelligent diagnostic algorithm. In this way, GenAI satisfies the needs of the general public for high-quality medical resources at broader, deeper and more popular levels.

Tencent created an interdisciplinary research team to boost the explainability of AI in the medical field. CT Image-assisted Pneumonia Triage and Evaluation Software, a descriptive document and operating manual developed by Tencent Miying, clearly defines user qualifications and security levels, provides a detailed description of the working principle of the product for AI professionals, and includes a detailed analysis on the sources, quantities, and multi-dimensional distribution of the training and test data to help AI professionals and product users (such as doctors) understand the model characteristics behind the software, provide the transparency and explainability of medical AI models, and clear concerns about the output deviation of the models caused by bias in the training data.

Ant Group independently develops a GenAI system to assist in diagnosis and treatment of pets. Through interactions with pet owners, this system can automatically collect the symptoms and behavioral characteristics of pets and accu-

rately diagnose common diseases in pets with the help of massive pet medical datasets and large model algorithms. It can also answer users’ questions about suggestions on pet raising, provide pet owners with a convenient health consultation channel and professional advice and help.

Quark Health Assistant provides professional and accurate health advice to promote medical inclusion. Quark utilizes generative AI technology to provide medical information retrieval services, where users can get further medical information and health advice through multiple rounds of human-computer interaction through characteristic descriptions of physical symptoms, self-diagnosis pre-diagnosis, and other features. At the same time, Quark combines its own search technology accumulation with knowledge enhancement and evidence-based retrieval to improve the quality of generative AI’s answers, with accuracy, relevance, and logic surpassing that of ordinary medical information searches.

(7) Cultural Life Enriched by GenAI + Entertainment

IBM adds AI-enabled digital experience to the Wimbledon and US Championships. The AI-enabled commentary feature developed IBM makes use of GenAI capabilities to generate voice-over and commentary containing different sentence structures and specific vocabulary. This makes the edited video content more informative, provides fans who watch event highlights in videos with more insights, helps fans capture critical moments of the event, and improves viewers’ experience.

Weibo uses GenAI to launch “AI Chat” that covers multiple verticals. By making use of the expertise of bloggers in different vertical fields and the stylistic and professional chat interac-

tions, “AI Chat” provides users with information value, optimizes their social experience, and improves the interactive stickiness between fans and bloggers. At present, this technological feature has been implemented in a series of highly professional fields, including emotions (horoscopes), law, finance, etc.

Tencent’s AI technology team promotes the positive applications of deep synthesis and content generation technologies. The team strengthens ethical governance from the perspective of countermeasure technology. They develop the face synthesis detection technology to achieve detection of photos and video synthesis and editing. “FaceIn Anti-fake Recognition”, a face synthesis detection platform built by the team, supports the detection of different face changing methods. It ensures the safe application of GenAI in the entertainment field by detecting and recognizing AI-generated and synthetic contents.

Baidu has launched the AI art and creativity auxiliary platform “Wenxin Yige” relying on PaddlePaddle and ERNIE. By providing a variety of AI image generation services such as image editing, skeleton and line draft recognition, small-sample training, etc. It targets individuals with design needs and creative ideas, intelligently generating diverse AI creative images. “Wenxin Yige” not only inspires visual content creators such as artists and designers, assisting artistic creation, but also provides high-quality illustrations for writers and other text content creators efficiently. Users and corporate clients can experience industry-leading text-to-image generation capabilities on Wenxin Yige.

(8) Cost Reduction and Efficiency Improvement Driven by GenAI + E-commerce/Customer Services

Amazon uses GenAI to provide personalized product contents for customers. Digitile, a

generative e-commerce SaaS solution provider, is using GenAI to improve the quality of product descriptions for its client Amazon, with an aim to improve discoverability and conversion rates. Compared with humans, AI-driven natural language processing can generate more accurate and detailed product descriptions and optimize the descriptions in product lists. Thus, it can improve customer satisfaction, reduce returns, save the operating costs of merchants.

JD.com launched a virtual anchor to help the platform with cost reduction and efficiency improvement. ChatRhino (Yanxi in Chinese), the virtual anchor developed by JD.com, has been widely used, which reduces the cost of live streaming for the platform by 95%, increases the average GMV by more than 30%, and causes GMV to increase by millions on a daily basis.

Xiaolce develops a virtual anchor to empower cross-border e-commerce live streaming. The virtual anchor developed by Xiaolce is embedded with different language types under the same image. As a result, the single anchor can live broadcast in different languages. Now this virtual anchor has appeared in cross-border live streaming on TikTok.

(9) Search Solutions Optimized by GenAI + Search

Google creates the GenAI-enabled search experience. Google unveiled a feature called Search Generative Experience (SGE). This feature embeds GenAI technologies in search engines and supports dialogue and text generation to improve the search experience of users.

Baidu has developed Wenxin Baizhong, an industry-level search system that empowers search engines. Wenxin Baizhong replaces the complex features and system logic of traditional search engines with simplified strategies and system solutions and provides access to various

enterprise and developer applications at a low cost. With a data-driving model, it optimizes industry efficiency and improves application effects.

360 launched a new search engine AI search, empowered search innovation search. 360 Smart Brain x 360 search engine, to create a new generation of intelligent search engine tools. AI search can generate relevant responses or content based on the text entered by the user, based on the understanding of the user's intentions and emotions, but also associations and creativity, with their own subjective views and styles. Compared with the traditional search engine based on keyword matching, it is more like an intelligent partner that can talk, think and innovate.

(10) Work Efficiency Improved by GenAI + Office

Microsoft launched Microsoft 365 Copilot to improve work efficiency. AI capabilities are integrated into Microsoft 365 to provide users with supportive capabilities in workplace and at home. Therefore, users can perform daily routine operations faster and more creatively and use AI in a responsible way.

Baidu launched an enterprise-level knowledge management platform called Infoflow to improve work efficiency. Based on Baidu's large language model ERNIE Bot, this platform has enabled a series of applications such as intelligent authoring plug-ins and intelligent meeting minutes. While browsing and using this knowledge base, users can easily call ERNIE Bot to automatically generate various types of content, thereby working more conveniently and efficiently.

Yinxiang Biji optimizes products and services through "EverAI". Yinxiang Biji launched "EverAI" service, which utilizes generative arti-

cial intelligence to provide its series of productivity tools with intelligent capabilities such as writing assistant, document analysis, and private note conversation to improve the efficiency of information processing and knowledge management. At the same time, it empowers smart hardware such as e-ink office notebooks and conference headsets to offer intelligent reading, writing, translation, meeting summarization and other services. "EverAI" is also deeply integrated into Ever Inc.'s enterprise service, Yinxiang TEAMS, to support teams and enterprises in many knowledge-intensive industries, such as finance, technology, education, healthcare, media and other knowledge management.

Cisco introduced a GenAI assistant to enhance security capabilities. The GenAI-driven Security Operations Center (SOC) created by Cisco provides comprehensive situational analysis for security analysts, orchestrates intelligence on Cisco Security Cloud Platform solutions, and provides actionable suggestions.

Alibaba uses GenAI to empower accessible work. Ding Talk and a number of companies jointly founded the "Smart Office Hardware Accessibility Alliance", in an attempt to explore information accessibility construction in the office environment and help disabled people and vulnerable groups participate in social affairs equally and seek equal job opportunities. In addition, drawing on a series of supportive policies and green channels launched on Alibaba's e-commerce platform to help disabled people address employment problems, DingTalk provides greater assistance in a way that "gives the needed what they need and not what they deserve" and helps hearing impaired people work decently.

360 AI digital employee empowers enterprises to improve efficiency and promote enterprise digital intelligence office. 360 AI digital employ-

ee is a generative AI with content generation, content understanding, and logical reasoning capabilities based on the 360 Smart Brain. By default, the AI digital employee provides eight conversational AI assistants that can help enterprise employees complete tasks such as copywriting, document analysis, planning and translation. At the same time, the platform also provides application modules for basic scenarios such as AI marketing creativity, AI clerical writing, AI document analysis, and AI translation. As a new paradigm of human-machine collaboration, AI Digital Employee will improve the productivity of enterprises and liberate people from repetitive labor. At present, 360 AI digital employees have been applied to office writing, marketing creativity, report analysis, knowledge quiz and other scenes, empowering a hundred lines of work and thousands of industries, helping enterprises reduce costs and increase efficiency.

(11) Smart City Constructed by GenAI + Urban Management

Baidu developed the large smart city model to assist in urban management. In 2022, Harbin, known as the ice city, worked with Baidu to develop a large smart city model that integrated urban development, AI computing power, algorithms and data. Harbin - Baidu·Wenxin integrates data knowledge across business, structures and departments in the city with multiple task algorithms to become a unified pre-training model based on ERNIE 3.0, an NLP large model developed by Baidu. The integrated model provides capabilities such as language comprehension and semantic analysis.

SenseTime empowers the reconstruction of urban space. SenseTime launched SenseSpace, a high-precision 3D scenario generation platform, to implement high-precision digital reconstruction of city-level large space for all AR applications of Hangzhou Asian Games. This

platform enables centimeter-level AR positioning of landscape inside and outside urban streets and venues while presenting accurate and smooth virtual-real integration effects. The Neural Radiance Field (NeRF) technology is used to automatically reconstruct 3D scenarios with details and lighting effects as vivid as real-world landscape, and leading AI+AR technologies are used to expose audience and athletes to various innovative applications.

Huawei Cloud Pangu Government Model allows for city event discovery in seconds and intelligent dispatch in minutes.

Huawei Cloud worked with the Shenzhen Futian district government to develop and launch the intelligent government service assistant Xiaofu, which has mastered a wealth of knowledge, such as administrative regulations and service processes and takes care of problem breakdown, multi-intent understanding, and government policy association. Through multimodal converged training combining natural language processing (NLP) and computer vision (CV), Pangu Government Model can dynamically parse city videos and images, enabling a closed loop for intelligent event handling that includes perception, understanding, handling, and decision-making.

(12) User Experience Optimized by GenAI + Information Services

SAP embeds GenAI into SAP solutions to optimize user experience. SAP plans to embed IBM's GenAI technologies into SAP solutions, in order to provide new AI-driven insights and automation, support and accelerate innovation, and create more efficient and powerful user experience. Meanwhile, relying on the new AI capabilities, SAP Start helps users leverage IBM's Watson AI solutions, which are constructed under IBM's Principles for Trust and Transparency and Data Privacy. Users will be able to improve their productivity under the support of the natural

language capabilities and predictive insights.

(13) Enhancing end-side big model applications by GenAI + Hardware Innovation

Arm China promotes the scale-up of end-side big models with the new generation of Zhouyi NPU. As generative artificial intelligence is hot, Arm China launched a new generation of programmable, highly parallel, flexible and scalable architectural design of the Zhouyi NPU and full-stack artificial intelligence software solutions, taking into account higher accuracy and flexibility, can support multi-core computing units, up to 320TOPS arithmetic sub-systems, and through the TSM task scheduling to give full play to the performance of the computing unit and i-Tiling technology to significantly reduce bandwidth, thus better supporting the generative artificial intelligence of the big model scale applications. The TSM task scheduling technology significantly reduces bandwidth to better support Transformer, the infrastructure for generative AI.

(14) Scientific Community Constructed by GenAI + Science Exploration

Zhipu AI utilizes generative AI technology to empower AI for Science. Based on its self-developed GLM-130B large language model, Zhipu AI has developed ChatPaper, which is a conversational private knowledge base that integrates searching, reading, and Q&A, and can be used as a research assistant to help scientific researchers improve the efficiency of searching and reading papers, as well as obtaining the latest research trends in the field.

Beijing Academy of Artificial Intelligence (BAAI) has built the “OpenComplex” platform and continues to delve into the field of protein structure prediction. As an open-source artificial intelligence algorithm platform for biological macromolecules, OpenComplex has already

open-sourced highly predictive structure prediction training and evaluation codes for proteins, RNA, and complexes. OpenComplex has established an end-to-end deep learning framework for three-dimensional structure prediction of biological macromolecules that unifies three tasks: “protein structure prediction”, “RNA structure prediction” and “protein-RNA complex structure prediction”.

Huawei Cloud Pangu Drug Molecule Model empowers new drug discovery and accelerates drug design. Huawei Cloud worked with the Shanghai Institute of Materia Medica of the Chinese Academy of Sciences to build this model, which was pre-trained on the data of 1.7 billion real drug-like molecules. A library of 100 million small drug-like molecules with 100% novel structures was generated. With this model, Professor Liu Bing, from the First Affiliated Hospital of the Medical School at Xi'an Jiaotong University, and his team quickly developed a new broad-spectrum antimicrobial drug called Drug X (cinnamobactin). This model accelerated the discovery of lead compounds, reducing the time needed from years to just one month. In addition, this model has helped six pharmaceutical companies and research institutes find new active small molecular compounds in antibacterial, anticancer, and cardiovascular and cerebrovascular fields.

Baidu explores GenAI applications in biological computing. Baidu has developed large biological computing models such as the compound characterization learning model HelixGEM and the protein structure prediction model HelixFold. HelixGEM-2 is a model that considers the multi-body and long-range interactions between atoms, and it can promote the research work in dual scenarios of the prediction of the chemical properties of quantum and virtual screening. HelixFold-Single, a second-level protein structure prediction model, is an open-source large protein structure prediction model based on a

single-sequence language model. It can improve the prediction results in the scenario of antibody structure prediction.

Huawei Cloud Pangu Weather Model helps people accurately obtain meteorological information and reduce losses. In July 2023, a paper titled “Accurate medium-range global weather forecasting with 3D neural networks” was published in Nature, reporting Pangu-Weather research results of the Huawei Cloud Pangu model R&D team. Pangu-Weather is the first AI model to have outperformed conventional numerical weather prediction (NWP) methods in terms of accuracy and speed (10,000 times faster). This model is able to forecast not only weather but also ocean waves, typhoon paths, and cold or heat waves. The European Centre for Medium-Range Weather Forecasts (ECMWF) and China Meteorological Administration (CMA) have started experimental use of Pangu-Weather.

NASA makes a joint commitment to earth exploration with IBM. NASA and IBM jointly released the largest open-source geographical spatial AI foundation models on the platform of Hugging Face. The purpose is to expand access to and application of AI technologies in climate and earth science research and thus to accelerate innovation and contribute to the construction of a more open, inclusive, and collaborative scientific community.

iFLYTEK has launched a large language model of scientific literature services to accelerate scientific research. iFLYTEK, together with the National Science Library of the Chinese Academy of Sciences, has jointly built a large language model of scientific literature service to provide multiple capabilities, such as research survey, scientific knowledge acquisition, paper reading, literature analysis, scientific writing, experiment design and so forth, so as to accelerate the efficiency of scientific research.



世界互联网大会
World Internet
Conference