

REVIEW ARTICLE

Venturing into the "uninhabited zone" to explore the path of artificial intelligence

Bo Zhang*

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

ABSTRACT

As a significant breakthrough in the field of natural language processing, Chat Generative Pre-trained Transformer's (ChatGPT's) powerful language generation and transfer capabilities have not only changed the way of human-computer interaction, but also had a revolutionary impact on multiple industries. Currently, to develop the third generation of artificial intelligence (AI), issues related to knowledge, data, algorithm innovation, computational resources, and ethics need to be addressed. The Chinese model is mainly application-driven, but the basic theoretical research is relatively weak. The development history of AI shows that realizing artificial general intelligence still faces enormous challenges. In the upsurge of AI, we should remain calm and scrutinize its potential risks. China should increase the intensity of basic theoretical research, attract global talent, and jointly promote the healthy and sustainable development of the field of AI.

Key words: artificial Intelligence, deep learning, large language model

INTRODUCTION

People usually consider artificial intelligence (AI) as the use of machines, especially computers, to simulate human intelligence. However, this view is not comprehensive. Our understanding of the essence of human intelligence is still superficial, and there is still no unified standard for the definition of "intelligence" worldwide. In 2019, German-American neurophysiologist Koch admitted "We don't even understand the brain of a worm",^[1] let alone the complex human brain. Therefore, the precise connotation of the concept of "intelligence" is still unclear, which brings considerable challenges to the in-depth exploration and research in the field of AI.

The origin of AI can be traced back to the 1940s and 1950s when computer science was just emerging. Scientists, with infinite expectations for the potential of computers, began their exploratory journey. In the budding stage of AI, research mainly focused on basic

algorithms and simple logic. Scientists tried to endow machines with basic computational and logical judgment abilities, hoping that machines could think like humans. With the surging advancement of technology, especially the improvement of computer computing power, a series of models and algorithms such as artificial neural networks and deep learning have emerged, injecting new vitality into the field of AI and pushing it to unprecedented heights. Globally, the United States and Europe are undoubtedly the leaders in the development of AI. In particular, Silicon Valley in the United States has become the birthplace of the global AI industry, bringing together tech giants such as Google, Apple, and Microsoft. Their research and innovation always stand at the forefront of the industry, constantly expanding the boundaries of AI. From the early ELIZA dialogue system to the widely used Siri and Alexa today, AI is no longer a self-appreciation in the laboratory. At the same time, cutting-edge technologies such as machine vision and autonomous driving are gradually maturing. They

***Corresponding Author:**

Zhang Bo, Department of Computer Science and Technology, Tsinghua University, No. 30, Shuangqing Road, Haidian District, Beijing 100084, China. Email: dcszb@mail.tsinghua.edu.cn; <https://orcid.org/0009-0000-3610-3595>

Received: 14 June 2024; Revised: 17 June 2024; Accepted: 16 August 2024
<https://doi.org/10.54844/eer.2024.0628>

 This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which allows others to copy and redistribute the material in any medium or format non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

are not only widely used in industrial production but also gradually infiltrating all areas of our daily lives. Against the backdrop of the international development of AI, countries have regarded it as a national strategy, increasing investment, and striving to occupy a favorable position in competition. The United States maintains a leading position with its deep foundation in science, technology and capital. European countries try to enhance their own R&D capabilities through strengthening international cooperation. Japan, with its exquisite skills and creative spirit, has achieved remarkable achievements in areas such as robotics. In China, AI has become part of the national strategy. The government, enterprises, and academia are working together to promote the research and application of AI. Chinese tech giants such as Alibaba, Tencent, and Baidu have achieved significant achievements in cloud computing, big data, machine learning, and other areas. The 2024 government work report proposed: "Deepen the R&D and application of big data, AI, etc., carry out the 'Artificial Intelligence +' action, and build a digital industry cluster with international competitiveness".^[2] AI is an important driving force for the new round of technological revolution and industrial transformation. General Secretary Xi Jinping also pointed out that we should promote the "deep integration of AI with economic and social development, and promote the healthy development of our country's new generation of AI". "Artificial Intelligence" has thus become a hot word frequently mentioned by representatives and committee members during the Two Sessions this year.

In some sub-fields of AI, China has reached an internationally advanced level. However, when viewed from a global perspective, there is still a significant gap between the overall level of AI in China and the international advanced level. This gap mainly manifests in insufficient innovation capabilities. If not remedied in a timely manner, it could quickly transform into a gap in productivity and economic strength. This situation not only concerns the current competition in economic strength but will also profoundly affect the country's development trajectory, future prospects, and destiny. China currently faces multiple challenges, including insufficient talent reserves, the need to enhance technological innovation capabilities, and key issues such as data security. In the field of AI, related research is still in the initial stage, and there is still a long way to go in the future. Especially as an emerging discipline of multidisciplinary integration, AI needs to achieve deeper breakthroughs and progress at the theoretical and practical levels. We need to continuously strengthen talent cultivation, enhance technological innovation capabilities, and strengthen data security management to ensure the healthy and orderly development of AI technology.

TWO PATHS OF AI

Materialistic school

After years of exploration and practice, AI has formed two main development paths. One is widely known as the behaviorist path or behaviorist school, sometimes also referred to as the materialist school. The core view of this school is to simulate and replicate human intelligent behavior through machines. It needs to be clarified that "intelligence" and "intelligent behavior" are two completely different concepts (Figure 1).

"Intelligence" refers to the complex processes that occur within the human brain, and our understanding and knowledge of "intelligence" itself is still quite limited. "Intelligent behavior", on the other hand, is the external manifestation of intelligence, which we can directly observe and simulate. Therefore, the development goal of AI mainly focuses on making the behavior of machines as close as possible to or imitate human behavior, rather than pursuing the complete consistency of internal working principles.

Take the chatbot program Chat Generative Pre-trained Transformer (ChatGPT) developed by Open AI as an example. It has successfully achieved this goal. The conversation experience with ChatGPT is very similar to a real human conversation, but internally, its working principle is not the same as the working principle of the human brain. This fully demonstrates that the current AI is on a path of machine intelligence, which is not exactly the same as human intelligence in essence, but only achieves a high degree of similarity in behavior. This is also the mainstream development direction in the field of AI today.

Idealistic school

In the vast field of AI, another important academic faction has emerged, named the innatist or idealist school. This school firmly believes that only by accurately simulating the working mechanism of the human brain can we achieve "intelligence" in the true sense and thus reach the ultimate goal of intelligence. This school is sometimes referred to as brain-like computing.

It's worth noting that these two schools are not absolutely opposed. They each explore in-depth the development path of AI based on different philosophies and methods. It's worth noting that these two schools are not absolutely opposed to each other. The behaviorist school argues that besides human intelligence, machines or other methods can also pave the way to intelligence. The innatist or idealist school, on the other hand, firmly believes that simulating the working principles of the human brain is the only way to

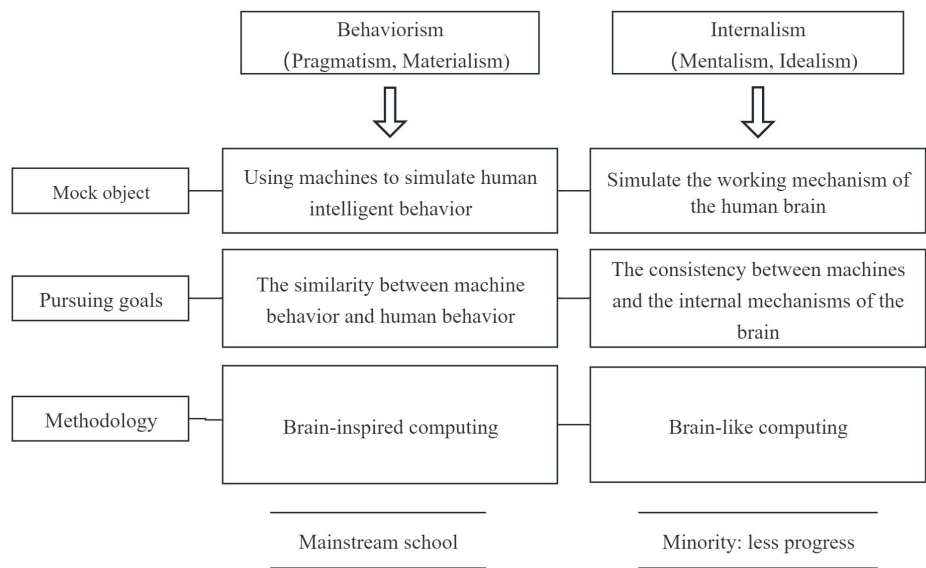


Figure 1. The path to realize AI. AI: artificial intelligence.

achieve intelligence.

Despite the different philosophies, both schools are currently in a stage of continuous exploration and trial. Tracing back to 1956, the field of AI officially kicked off. At that time, a conference on AI was held in the United States, where ten outstanding scholars from various fields such as mathematics, computer science, cognitive psychology, economics, and philosophy gathered. After eight weeks of in-depth discussion, they jointly defined the basic concept of AI and clarified its development goal—to create a machine capable of thinking like a human. To achieve this goal, the participants proposed the method of using symbolic reasoning and symbolic representation.

At this historically significant seminar, Newell and Simon presented a program called "Logic Theorist". This program successfully proved some theorems in Chapter 2 of Whitehead and Russell's "Principia Mathematica" through the machine, which not only proved the machine's ability to reason, but also further confirmed the machine's potential in logical reasoning. This conference not only laid a solid foundation for the development of the field of AI, but also pointed out the direction for subsequent research. At the same time, it provided an important reference for defining the concept of "AI".

THE THREE STAGES OF AI DEVELOPMENT

As a field, the history of AI has been relatively young

since its inception in 1956, less than 70 years ago. In its evolution, it can be divided into three stages: first-generation AI, second-generation AI, and third-generation AI.

First-generation AI

The original intention of the first-generation AI researchers was to build machines capable of human-like thinking, that is, to design a machine that can think. Thinking refers to reasoning, decision-making, diagnosis, design, planning, creation, learning, *etc.*, which essentially form the foundation of our white-collar work. Any white-collar job cannot be separated from the application of this thinking ability or rational behavior. In the field of medical diagnosis, doctors' thinking ability usually far exceeds that of ordinary people. The main difference between doctors and ordinary people lies in the fact that doctors have rich medical knowledge and clinical experience, while ordinary people lack these foundations and cannot diagnose and treat; in addition, besides having medical knowledge and clinical experience, it is also necessary to have the ability to apply knowledge. This ability is mainly manifested as reasoning, that is, being able to analyze and judge from one thing to another, from the surface to the inside, based on existing knowledge, thereby drawing new conclusions and acquiring new knowledge.

Whether engaged in management work or technical work, the abilities required can mainly be divided into two aspects. One is to have rich knowledge and experience in a certain field, and the other is to have strong reasoning ability. Based on this analysis, the

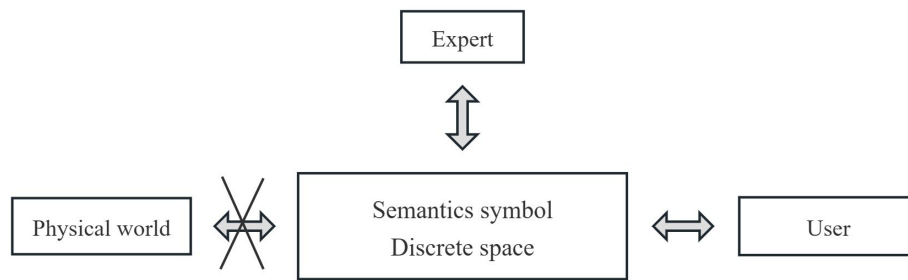


Figure 2. Limitations of the knowledge-driven models.

founders of AI proposed the concept of a "reasoning model based on knowledge and experience". In short, to realize the thinking ability of a machine, humans only need to integrate the corresponding knowledge into the computer system. For example, if people want the computer to diagnose diseases, they only need to integrate the doctor's professional knowledge and clinical experience into the knowledge base, and embed the doctor's reasoning process of diagnosing diseases into the reasoning mechanism, then the computer can simulate the doctor to diagnose diseases. If you want the computer to take on thinking work such as design, humans need to input the design knowledge and experience of a certain field into the knowledge base, and at the same time, embed the reasoning process of design into the reasoning mechanism, then the computer can assist humans in design work. Therefore, the reasoning model based on knowledge and experience is a common computational model for all rational behaviors. With this computational model, machines can have thinking abilities similar to humans.

The core idea of this model is knowledge-driven, in other words, knowledge is the source of human wisdom, and knowledge is power. This knowledge is usually called rational knowledge, and rational knowledge comes from learning. Therefore, all those engaged in white-collar work must receive education. Because without education, you cannot get rational knowledge, and without rational knowledge, you cannot do any white-collar work. Therefore, many people need to receive education, and especially pursue higher levels of education, to ensure that they can better complete work that requires rational thinking. Sometimes, this process can be viewed as symbolism, because it stores human knowledge and experience in symbolic form in the computer system.

In the early 1970s, American scholars at Stanford University first applied this idea and developed a medical diagnosis system called MYCIN, which was mainly used for the diagnosis of blood infections and antibiotic

prescriptions. The MYCIN system integrates the knowledge of internal medicine doctors and infectious disease experts. With these limited knowledge resources, the system can diagnose blood infections like an internal medicine doctor and prescribe antibiotics. At the same time, the system also integrates the professional knowledge of infectious disease experts, so it surpasses ordinary internal medicine doctors at the diagnostic level. Developed countries may need high-level medical diagnostic systems, but this is not necessarily required in developing countries, because the existence of community doctors can provide a certain degree of medical services, so the medical diagnostic system only needs to reach the level of doctors in tertiary hospitals to promote use. In addition, the system can also assist general practitioners in medical auxiliary diagnosis.

This system seems to have many advantages, because it can reason like humans, and humans can fully understand its diagnostic process, so this system is understandable and explainable. But the biggest drawback is that all knowledge has to be told to it by humans, and it is impossible to learn knowledge from the physical world (Figure 2). In the era of first-generation AI, computers do not have the ability to self-learning, which is undoubtedly a huge challenge. Because humans teach knowledge in natural language, however, in the process of knowledge transmission to the computer, facing a computer that did not have the ability to understand natural language at that time, a communication method that the computer can understand must be explored to build the corresponding system. Therefore, it was very difficult to construct such a system at that time, and it was time-consuming and laborious, so its application and industrialization were relatively poor. Since its source of knowledge completely depends on human input, in this case, it cannot surpass human wisdom.

The first-generation AI model is based on a highly difficult-to-construct knowledge and experience reasoning model. The knowledge scope of the MYCIN

medical diagnosis system developed by Americans is relatively limited, only limited to prescribing antibiotic drugs and diagnosing blood infections, but even such a functionally simple system took as long as three and a half years to build. So the development process of first-generation AI is quite tortuous, the initial application is limited, and after the expert system later, although there are some applications in narrow fields, overall, its application range is still relatively narrow. Therefore, that period was vividly called "Artificial Intelligence Winter" or "AI Winter" for short.

Second generation AI and its development in China

During the downturn of the first generation of AI, the second generation of AI emerged. The second generation of AI mainly started from artificial neural networks, which can be traced back to the artificial neural network model proposed in 1943. It primarily aimed to simulate the working principles of human brain neural networks. This model was initially very simple, and the tasks it could perform were relatively limited. The early progress of the second generation of AI was also slow.

The start of the second generation of AI in China: Tsinghua University's AI and intelligent control teaching and research group

Tsinghua University has been involved in the field of AI since 1978. For a relatively long period, it was in the stage between the end of the first generation of AI and the beginning of the second generation, encountering the "AI winter" (Figure 3).

In 1978, Tsinghua University established the Artificial Intelligence and Intelligent Control Teaching and Research Group, which was the earliest teaching and research institution for AI in China. At that time, there were about 30 teachers involved, most of whom came from the field of automatic control and were not originally engaged in AI research. The first batch of master's students was admitted in 1978. From 1980 to 1982, I visited the United States as a visiting scholar. During my visit, I found that the director of the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign was the famous Chinese American Professor Robert T. Chien, whose main research direction was AI. In his team, there were six doctoral students, one of whom had been studying for 8 years, but was unable to continue research due to the inability to select an appropriate research topic; another doctoral student switched to other directions in computer science after less than a year of research; and another doctoral student from Taiwan, China despite having studied for four years, also failed to find a suitable research direction. From these situations, it can be seen that AI was indeed in a trough at that time.

Tsinghua University began to admit the first batch of doctoral students in the field of AI in 1985. Although it was possible to carry out teaching related to AI, it was difficult to conduct scientific research. Therefore, from 1982 to 1984, we conducted in-depth investigations. With the support of the Ministry of Ordnance Industry, we visited numerous research institutes and factories related to ordnance in the southwest and northeast regions. After this investigation, we deeply felt the necessity of developing intelligent robots. Because at that time, foreign countries had already automated the areas of artillery shell assembly and explosive handling, while in China it still relied on manual operations. Therefore, at that time, we considered intelligent robots as a major research direction.

Based on this, we began to establish an intelligent robot laboratory, which faced many challenges at the time. The primary problem was the lack of funds. The original plan was to purchase the PUMA560 robot, because in all universities in the United States engaged in robotics research, this equipment was standard. However, because the robot was listed as a banned material for China by the Paris Coordinating Committee, we could not purchase this key equipment at that time. Through cooperation with Fujian Province, we purchased second-hand equipment from Hong Kong and shipped it back to China in the name of drilling machines. Under very difficult circumstances, we successfully purchased the first imported robot equipment in the country, which came without any instructions. This was the first robot imported nationwide. At that time, the equipment was worth hundreds of thousands of yuan, and we lacked sufficient funding. Therefore, we formed a partnership with the Fujian Provincial Computer Research Institute, proposing that each party provide half of the funds, and borrow the other half of the funds from the institute. In effect, the institute provided all the funds to purchase the equipment. Later, because our work achieved significant results, the institute never asked us to repay that part of the loan. In this way, we successfully established this laboratory.

Tsinghua University established the laboratory in 1985, and then in 1986, the state established the "863" High Technology Development Plan, which took intelligent robots as a theme. Since then, Tsinghua University has actively participated in the first "863" high-tech research on the theme of intelligent robots. From the first to the fourth, Tsinghua University teachers participated in the work of the relevant theme expert group as experts. By the fifth, Tsinghua University became the leader in conducting research on intelligent robots. In 1997, space robots were established, and Tsinghua University was also the leading unit for space robots. Under these circumstances, we began preparations in 1987 and formally established the "Intelligent Technology and

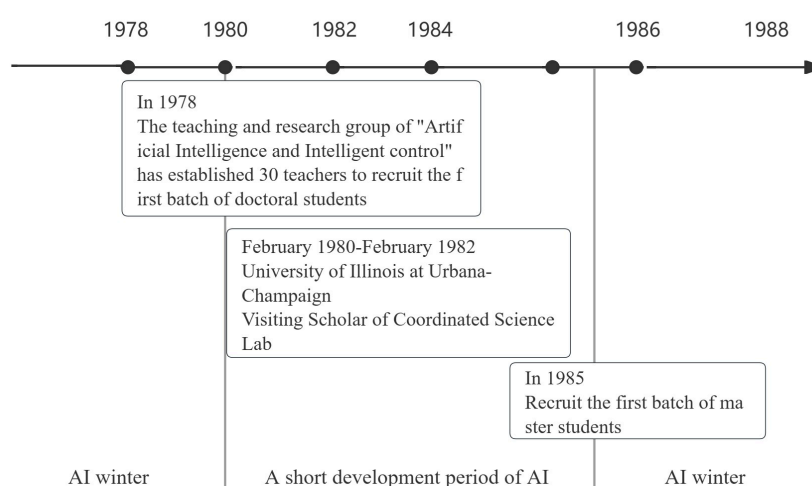


Figure 3. AI start of Tsinghua University: Planning and exploration of robots and robots. AI: artificial intelligence.

Systems" State Key Laboratory in 1990. The laboratory has achieved remarkable results, with evaluations conducted every 4 to 5 years, and it has received excellent ratings three times in a row. Thanks to these excellent evaluation results, Tsinghua University received up to 10 million in operating funds, which was a considerable amount of money at the time. It was due to the strong support of these two tasks that the school's research work was able to proceed smoothly. During the downturn, many units, both at home and abroad, faced difficulties in continuing to advance related research. However, from 1978s to the late 1990s, due to the support of the State Key Laboratory for Tsinghua University, our research work has always maintained a good development momentum. Especially in theoretical research work such as search, planning, and problem-solving, significant progress has been made with considerable international influence.

Since 1990, this research team has been dedicated to the research of autonomous driving cars (then called mobile robots), which is one of the earliest units to carry out this work internationally, and is even more in a leading position in China. In 1992, the research team further undertook the research of a military mobile robot project, which later also received rewards from the state and relevant departments. This work has continued to this day. Although many units are now conducting research on autonomous driving, Tsinghua University is actually one of the earliest units to conduct research on autonomous driving.

Characteristics of the second generation of AI: Emphasis on deep learning and data-driven approaches

The early development of the second generation of AI faced a significant challenge. Due to the relative simplicity of the models and the fact that many effective learning algorithms had not yet been discovered, initial progress was relatively slow. However, since the beginning of this century, the second generation of AI has flourished.

The first generation of AI was primarily guided by symbolism, aiming to simulate human rational behavior. However, human behavior is not only limited to the rational level, it also includes a rich array of perceptual behaviors. These perceptual behaviors are simulated using artificial neural networks.

As previously discussed, knowledge plays a crucial role as the source of human wisdom. It serves as the basis for our rational behavior. This knowledge refers to rational knowledge such as methods of problem-solving, which primarily come from education. In addition to rational knowledge, humans also possess a vast amount of perceptual knowledge.

For example, the understanding of objects like horses, cows, sheep, or individuals like "Zhang San" and "Li Si" belongs to perceptual knowledge. The acquisition of perceptual knowledge does not originate from book learning or oral instruction. When you try to explain what a "horse" is in natural language, you might say that a horse has a horse's head, a tail, and four legs. But to clarify the concepts of "four", "legs", *etc.*, it would involve many more concepts. How to accurately describe the attributes such as "long and thin" and "four", becomes a significant challenge that cannot be ignored. In other words, if you want to convey a

perceptual concept, you usually need to use multiple new concepts to describe it. However, how to convey that initial perceptual concept is a challenging problem. Therefore, perceptual knowledge cannot be taught through language and cannot be learned from books. The first perceptual knowledge everyone obtains is the cognition of the mother, but when and how to recognize one's mother is still a difficult question to answer. If we can unravel the mystery of where perceptual knowledge comes from, we might find a way to teach computers to recognize things like horses, cows, and sheep.

The underlying logic of deep learning: Observation and listening

If you observe children closely, you will find that when they are very young (especially before the age of 2), apart from basic physiological needs such as eating, drinking, defecating, urinating, and sleeping, they also need to complete four crucial tasks. If these tasks are not effectively completed, the child's cognition, intellectual quotient (IQ), and emotional quotient (EQ) development will be greatly affected. The first task is observation. The child needs to use all the time to seriously observe the things around him. After waking up, children usually focus their attention on objects, and the reason for continuous gaze is to take full advantage of the time to observe the surrounding environment in detail to build a visual foundation. The second task is to listen, to establish an auditory foundation. Children always respond positively to sounds, and in this process, they need to continuously stimulate their hearing to help establish an auditory foundation. The acquisition of perceptual knowledge relies on continuous observation and listening, i.e., unsupervised learning. In the second generation of AI, deep learning uses this method.

In the past, humans usually used programming to inform computers about the features of horses, cows, and sheep. However, this method has obvious shortcomings, resulting in less than ideal results. The same situation also occurs in the field of speech recognition. People once tried to make computers understand "I" by teaching the features of its pronunciation, but the results were not satisfactory. To solve this problem, humans have finally found an effective solution based on machine learning technology with large-scale data. Humans have collected a large number of photos of horses, cows, and sheep on the Internet, most of which are used as training samples for observation and learning. After the learning process is completed, the remaining photo samples are used as test samples to verify the recognition effect, such as achieving a recognition rate of 95%, *etc.* The same method has also been adopted for speech recognition processing. Humans have collected a large amount of speech data, most of which are used as training samples

for the computer to listen and learn, and the remaining part is used as test samples to test the learning results of the computer. This is also the same method, namely observation and listening. So, what is used to observe and listen? It is the artificial neural network. In other words, the problem of recognition is treated as a classification problem, using artificial neural networks for classification. Specifically, data about horses is categorized as one class, and data about cows is categorized as another class, thereby realizing the recognition function. A multi-hidden layer (deep) neural network is usually adopted, and the method of learning with this type of deep neural network is called deep learning. It's important to note that deep learning does not refer to the depth of learning, but to the learning using neural networks that have depth. This method can usually achieve very good results.

Classifying, learning associations, and making predictions through deep learning

The effect of deep learning is shown in Figure 4. In the past, image recognition relied on the same image network (ImageNet). These resources cover more than 20,000 categories and include a total of 15 million images. Before that, humans used programming methods to define to the system what is the feature of a horse, a sheep, *etc.*, but this method led to a false recognition rate as high as 50%. Later, machine learning was used for image recognition, and the result was a significant reduction in the false recognition rate, down to 3.57%. This progress has reduced the false recognition rate from 50% to 3.57%, which even surpasses the false recognition rate of humans, with an average human false recognition rate of 5.1% (Figure 5).

In addition to using artificial neural networks for classification, they can also be used to discover the relationships between data. Current machine translation uses this principle. Traditional machine translation methods that use grammatical and semantic analysis have difficulty achieving good results, but the use of machine learning methods has proven to be significantly effective. Because it is about finding the relevant relation between Chinese (data) and English (data), rather than a detailed analysis of their grammar and semantics. By end-to-end learning a large amount of data, the model can quickly output the corresponding English translation based on the input Chinese. The underlying principle is that the model has successfully linked Chinese and English closely by learning the relevant relations in the data, thus achieving the function of translation. In addition, predictions can also be made using methods such as artificial neural networks or machine learning. For example, predictions in the fields of infectious diseases, products, stocks, *etc.*, all stem from using historical data to predict future changes. Therefore,

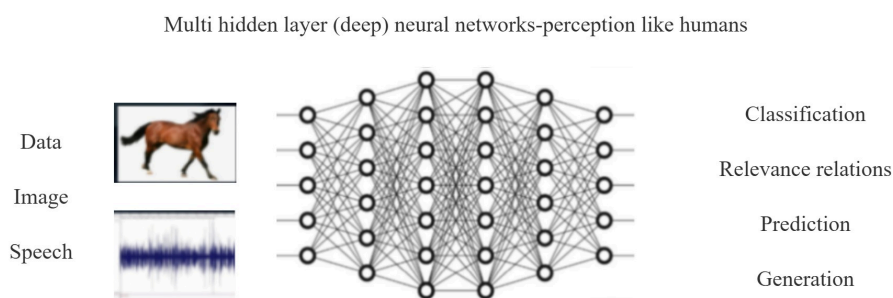


Figure 4. Deep learning.

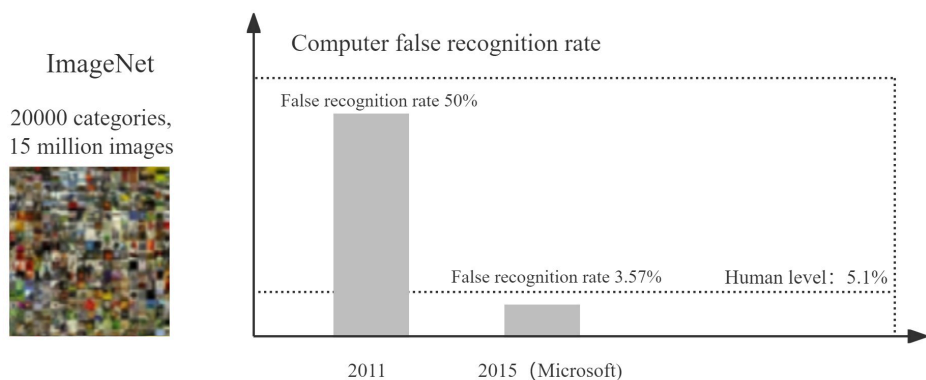


Figure 5. Classification: Image recognition.

machine learning technology in the era of big data has a wide range of application values. In addition to classification, learning associations, and prediction, it can also be used for content generation tasks (This will be discussed later).

When we entered the field of AI in 1978, artificial neural networks were still in their infancy, and research progressed slowly. But even in such a trough, we still made good contributions to the basic research of learning algorithms and model construction of artificial neural networks. The emergence of deep learning technology marks the prosperity and rise of the second generation of AI, which has also set off a research boom in China. Among them, Tsinghua University has emerged a large number of young participants, and the young team represented by Jun Zhu has done a lot of work in this area. In terms of theoretical work, they proposed a probabilistic learning theory and model. This theory adds a dimension to the original Bayesian theory, that is, the posterior distribution. The original Bayesian theory mainly relies on the prior distribution and the

likelihood function, and the introduction of the posterior distribution improves the quality and efficiency of learning. Based on this innovative theory, the AI basic theory research team successfully developed an open-source programming library that can be used worldwide. This work has also been widely recognized and praised.

Safety and reliability issues of the second generation of AI

All data (images, voices, etc.) of the second generation of AI come from the physical world, so it has certain advantage in actual applications, but it is difficult to elevate to the cognitive level. Its recognition function is limited to the distinction of objects, and it has not achieved real cognition of objects. This is its biggest problem, that is, it is unsafe, untrustworthy, uncontrollable, unreliable, and not easy to generalize.

One of our doctoral students discovered the insecurity of deep learning at an early stage, providing a very typical example. As shown in Figure 6, the snow mountain picture on the left, both the computer and

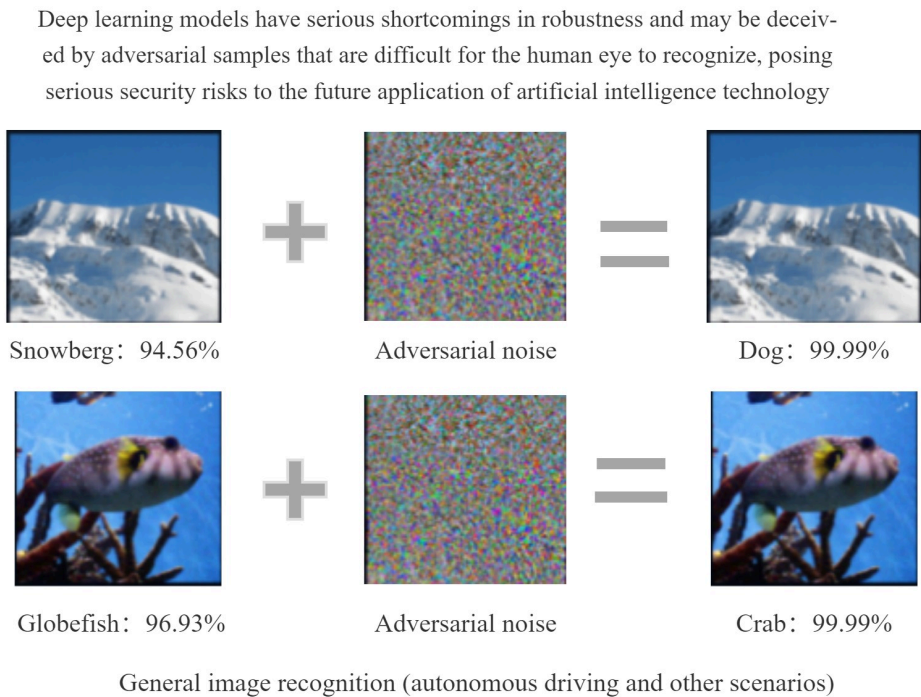


Figure 6. Security of the AI algorithm. Adapted from Dong *et al.*^[3] AI: artificial intelligence.

humans can recognize as a snow mountain, but with only a slight addition of noise, it becomes the picture on the right (the picture on the right has slightly more noise than the one on the left). To humans, it looks like a snow mountain, but the computer misidentifies the snow mountain as a dog. This fully demonstrates that pattern recognition based on deep learning is completely different from human vision. Although it can distinguish between snow mountains and dogs, it actually recognizes neither dogs nor snow mountains.

The question worth exploring in depth is—how to define a dog. Usually, it is distinguished by human vision, mainly by its appearance. The appearance of a dog is different from that of a cat, but the problem is that there are many types of dogs with different shapes. Why can humans distinguish which ones are dogs from the various appearances of dogs? Moreover, the same dog will have significant changes in appearance when standing, lying down, or running. It looks different from the front and the back. However, the human visual system can accurately identify dogs from various dog shapes. Why humans have this recognition ability is still not fully understood. There are still many deficiencies in human research in the field of brain science because we don't know why human eyes can determine that it is a dog in the ever-changing shape of a dog, even if sometimes part of the dog's body is hidden, humans can

still recognize it as a dog.

How human vision solves this problem is still an unsolved mystery, so there are still great challenges in computer simulation of human visual recognition capabilities. In the early stages, the computer could recognize a dog in a fixed position, but once the position of the dog changed, the computer could no longer recognize it. This reflects the computer's problem with dealing with displacement invariant, however, this problem has now been solved. For example, the computer performs well in recognizing dogs of a fixed size, but once the size of the dog changes, whether it is enlarged or reduced, the computer cannot accurately recognize it. This is called scale invariant. Currently, even with technological advances, computers still face difficulties in recognizing dogs with size changes, so now machines usually rely on local texture features to distinguish between dogs and snow mountains. Therefore, if the texture of a certain part of the snow mountain image is replaced with fur texture, and the shape remains unchanged, the computer will mistakenly recognize it as a dog, which is its essence. The above analysis shows that deep learning is still not safe, reliable, and trustworthy.

Early research has revealed security issues with facial recognition technology. As shown in the Figure 7, the

Deep learning models have serious deficiencies in their robustness, and will be deceived by some "adversary" samples that are not easy to identify by human eyes, bringing serious security risks to the future application of artificial intelligence technology

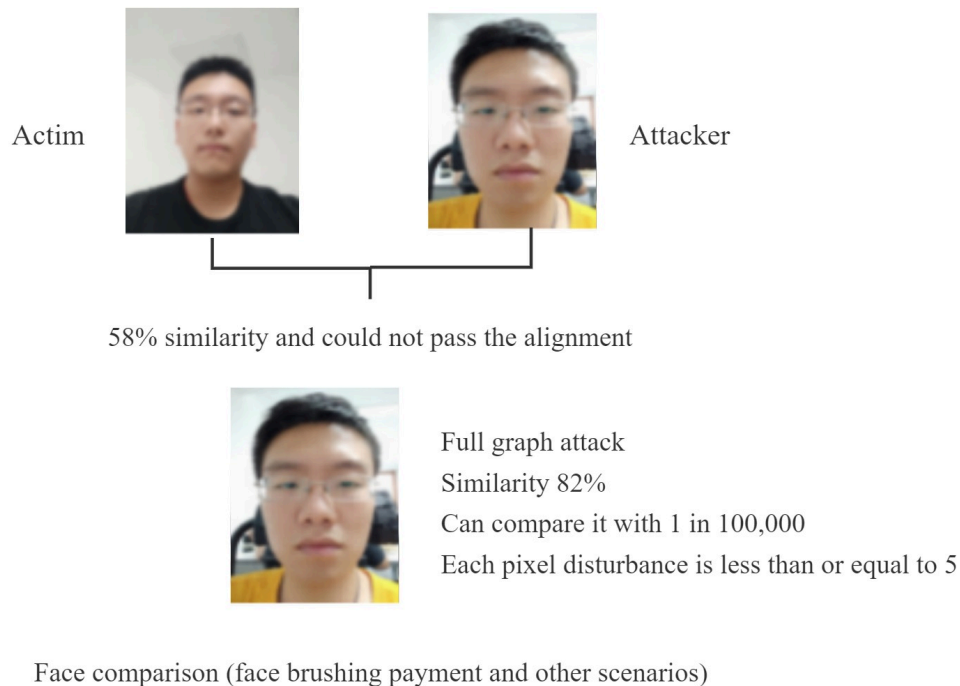


Figure 7. Security of the AI algorithm. AI: artificial intelligence.

computer can recognize that the person on the left and the person on the right are not the same person. However, by adding some slight noise to the image, the computer may mistake the person on the right for the person on the left. This is why facial recognition is considered a security risk. Because once you slightly change a person's local features, although humans can still recognize that this is the same person, the computer completely sees it as a different person. Therefore, the security of facial recognition technology is indeed a concern. Currently, using facial recognition technology to make payments may still be acceptable, but there are big problems if it is used to withdraw money. Online, people are not allowed to withdraw money just by swiping their faces, which further proves the insecurity of facial recognition technology.

According to CSRanking statistics, the team from Tsinghua University ranked first in the number of high-quality articles published in the world's top journals in the field of AI from 2017 to 2019.^[4] This shows that the average level of Tsinghua University and even China in this field is quite good, reaching an advanced level.

However, the problem is that there is still a big gap between our country's highest level and the world's top level, and it is difficult to cultivate truly outstanding talents. In other words, our country still needs to strengthen its innovation capabilities. The highest level is often a key factor in determining the development of science and technology, so our country still needs to continue to work hard in this area.

Third generation AI

Basic ideas to developing third generation AI

The basic approach to developing third-generation AI is that we must develop AI theory (Figure 8). Compared to the rapid and continuous development of information technology, the progress of AI seems relatively slow and full of twists and turns. Information technology has built a solid theoretical foundation from the beginning—the theory of Turing machines was established in 1936, and modern communication theory was established in 1948. Due to the establishment of these theoretical foundations, the development and application of technology can proceed smoothly. However, so far, the field of AI has not yet formed a unified theory, and the

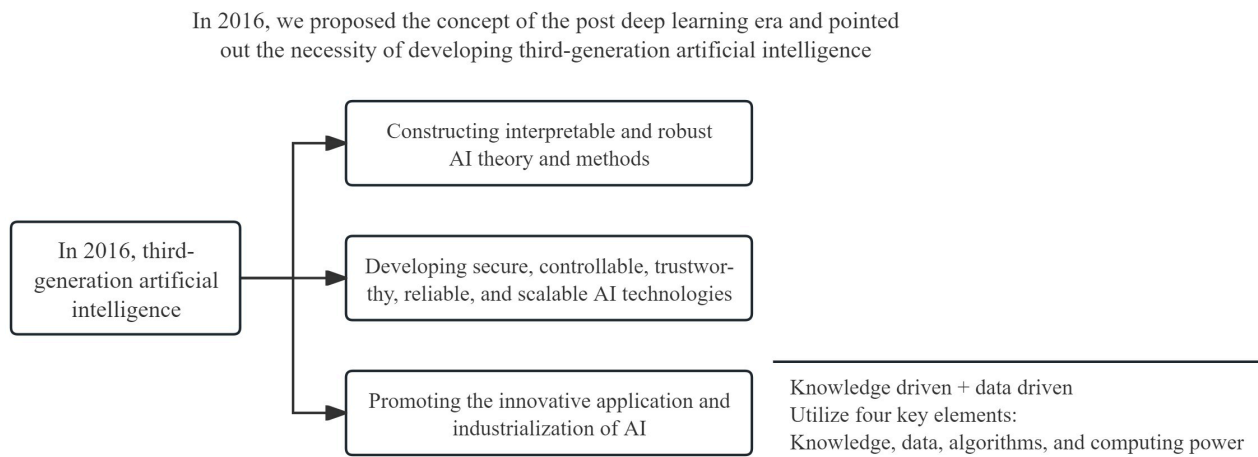


Figure 8. The basic idea of developing the third generation of AI. AI: artificial intelligence.

models and algorithms of the first generation of AI and the second generation of AI have many defects. Therefore, it is very necessary to develop the theory of AI, so that we can develop safe, controllable, trustworthy, reliable, and scalable AI technology.

If you need to apply AI in your work in the future, you must always be alert to possible security issues. Autonomous driving, intelligent banking and so on must pay special attention to safety issues, because for humans, when it comes to the handling of life and property, safety must be considered first when it comes to intelligence. For the current AI technology, the higher the intelligence level of the system, the more insecure it is. Many people mistakenly think that introducing AI technology can enhance security, but the fact is just the opposite, the widespread application of AI technology often increases security risks. Despite this, scientists are still committed to the development of AI technology, because it can significantly improve work efficiency and quality, of course, this also means that we must face and deal with the security issues that come with it.

The first generation of AI primarily utilized three elements (resources): knowledge, algorithms, and computational power, with knowledge being the most prominent. The second generation of AI mainly relies on data, that is, the three elements of data, algorithms, and computational power. Both generations of AI have their shortcomings because they only use the middle three elements. To overcome these limitations, we must simultaneously use knowledge, data, algorithms, and computational power. However, foreign perspectives emphasize the role of data, while we tend to focus more on the role of knowledge. Knowledge is the crystallization of human wisdom and cannot be replaced by

data. But at the same time, we should not ignore the importance of data, after all, computers far surpass humans in processing data. Our emphasis on data is actually emphasizing the role of machines; emphasizing knowledge is emphasizing the role of humans. The role of humans in AI should be more important than that of machines. The reason for the tremendous success of ChatGPT is precisely because it fully integrates the four key elements of "knowledge, data, algorithms, and computational power". The perfect combination of these four elements makes ChatGPT stand out in the field of AI and achieve remarkable achievements.

The two core elements of ChatGPT: Large models and large text

ChatGPT's powerful performance is mainly due to its two core elements: large models and large text.

Large model

Large Language Models, or LLM (Figure 9). Firstly, a "large model" refers to a massive artificial neural network. This network can be used not only for classification, and learning the associations in data, but also for prediction. Currently, ChatGPT is mainly used for language generation tasks. This massive artificial neural network is called a "transformer", which is the "T" in GPT. And the name GPT is actually composed of three parts: G stands for Generative, P stands for Pre-trained, and T stands for Transformer. These three elements together constitute the core technology of ChatGPT, enabling it to show excellent performance in the field of language processing. Deep neural networks are very deep networks that contain multiple layers, with their intermediate hidden layers typically consisting of several to tens of layers. The neural network of ChatGPT is very deep, with 96 layers. Humans have spent 74 years

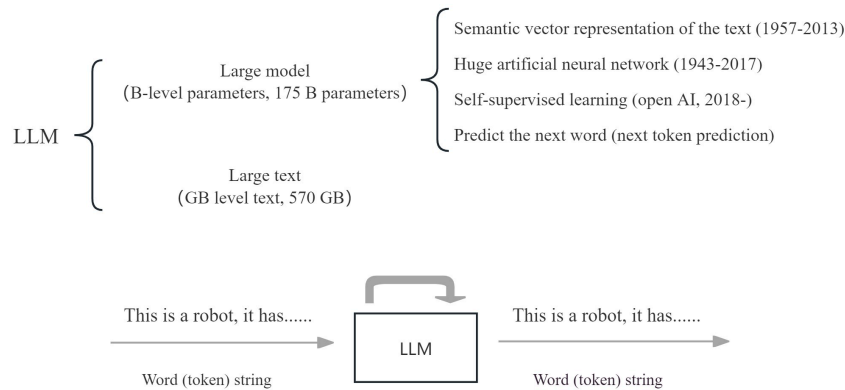


Figure 9. LLM. LLM: large language model.

solving the problem of attention-based massive artificial neural networks. This huge AI neural network is called a Transformer. With the Transformer, long texts can be input simultaneously.

Secondly, The neural network of ChatGPT also has a great width, capable of inputting more than 2000 Chinese characters (one token, roughly equivalent to one Chinese character) at a time. Previous neural networks input text character by character, but now they can input a large block of text at once. GPT4 Turbo can accommodate the simultaneous input of 128,000 tokens, that is, over 300 pages of text at once. It can see the entire text, which is the first "big". In text processing, a core issue is the semantic representation of text. In the past, text on computers was represented by symbols, which made it difficult for computers to understand their meanings directly. For example, in the sentence "I hit him", the computer only sees a series of symbols and does not know the specific meaning expressed by these symbols. Therefore, it is necessary to do syntax analysis for it, clarify that "I" is the subject, "hit" is the predicate, and "him" is the object, so that the computer can understand its meaning. Now, text is no longer represented by symbols, but is represented in the form of semantic vectors. This breakthrough was achieved by humans after 56 years of effort, from 1957 to 2013. This representation method means that words, sentences, or paragraphs can all be transformed into vector form, allowing computers to understand the meaning of the text directly through these vectors. Vectors with the same semantics will be placed in the same position, while vectors with different semantics will be distinguished in space. When the computer sees the vector "I hit him", it can immediately recognize its meaning because all vectors expressing similar meanings are clustered together. In the past, computers could only process text as data, but now they can treat it as

knowledge and process the content of the text. This is the most important breakthrough, the vector representation. Since vectors are essentially a series of values, computers have become more efficient and convenient in performing related calculations.

Finally, OpenAI contributed "self-supervised learning". ChatGPT is not just the result of four years of independent research and development by OpenAI, but the crystallization of continuous research and exploration by scientists and engineers around the world for six or seven decades. However, "self-supervised learning" was proposed by OpenAI. OpenAI's main contribution lies in proposing the "self-supervised learning method" and successfully applying it in practice, with the entire development process taking four years. In the past, to make machines learn, a lot of preprocessing and pre-labeling work was needed, which undoubtedly added a huge workload and limited the scale of machine learning. "Self-supervised learning" means that the original text can be learned without any processing, predicting the next word with the previous text, inputting the next prediction, and turning the next prediction into the input, and so on. This is a new learning paradigm.

Large text

After the advent of self-supervised learning, all text can be learned without any preprocessing. The amount of text data has also jumped from the original GB level to the TB level. At present, OpenAI has learned about 45 TB of data, equivalent to the capacity of 13.51 million Oxford dictionaries. And it doesn't just read mechanically, it understands its content, and reads its internal knowledge. This has brought humans into a new era, known as the era of generative AI. In this era, the most notable achievement is the powerful language generation ability shown by ChatGPT, and the powerful language

generation ability is shown in the open domain, not limited by the domain. First, when chatting with ChatGPT, you can discuss issues in all domains, which is undoubtedly a major progress in the field of AI. Whether it is the first generation of AI or the second generation of AI, its activities are limited to three conditions—completing specific tasks with specific models in specific domains. The "three specifics" are so-called narrow AI, and dedicated AI. ChatGPT completely breaks this problem, there is no domain limit when chatting with it. Secondly, ChatGPT can generate diverse outputs, which is also the soul of ChatGPT. This diversity not only gives it the possibility of innovation, but also gives it the fact that it may make mistakes. The more you hope it outputs creatively, the more you have to allow it to make mistakes. This also explains why sometimes ChatGPT's answers are exceptionally witty and smart, and sometimes they may say absurd words, which is the inevitable result of pursuing its diverse output.

The text generated by ChatGPT is semantically coherent, human-like text. This is quite surprising. It has learned a lot of text data, but it can be so orderly. If its output is incomprehensible or illogical content, it will be very tricky because it is impossible to accurately capture its real intention and it is difficult to carry out subsequent research work. Fortunately, the content expressed by the system at present is all clear and understandable human language, even if it is nonsense, it is at least language that people can understand. In this field, the reason why OpenAI is worthy of affirmation is mainly because few institutions dare to take risks and make such significant progress. OpenAI invested hundreds of millions of dollars to develop this system, although initially many people were skeptical and thought that its output results might be chaotic. However, it is surprising that the content output by the system is all smooth, semantically coherent, and human-like text, which is of great significance and highly commendable. The second major breakthrough is the realization of human-computer natural language dialogue in the open domain, and now the dialogue with ChatGPT is no longer limited to a specific domain. Before the advent of ChatGPT, it was generally believed that it would take several generations of relentless effort to achieve natural language dialogue with machines in the open domain. In the past, even in relatively narrow domains, it was impossible to have natural language dialogue with computers, but now it can break through domain restrictions for dialogue. These two breakthroughs are indeed eye-catching. It was originally expected that the development of AI would require the joint efforts of several generations, but surprisingly, such significant progress has been made in 2022, which is undoubtedly a very remarkable achievement in the field

of AI.

A test is conducted on Chat General Language Model (ChatGLM), a language model with hundreds of billions of parameters that has question-answering and dialogue capabilities, concerning an AI system's ability. The system is asked to write an essay from the perspective of a third-grade student, titled "The Flying House" (Figure 10). The key to this test is to judge whether the system successfully simulates the thinking perspective of a third-grade student during the writing process. Some systems perform poorly in this regard, often unconsciously incorporating adult language and concepts, turning the essay into an overly sophisticated science fiction story, which is obviously not in line with the requirements. The title "The Flying House" itself is very creative and imaginative, giving the system a great deal of room to play. Some excellent systems can create imaginative articles, depicting the house flying in the sky, passing through cities and rural areas, with everything in sight being fairy-tale-like magical scenery. Therefore, the LLM is a step towards general AI. Microsoft views the LLM as the spark of general AI, and the view of Western public opinion is that it is the dawn of general AI. However, a minority believe that it is general AI, which is a somewhat biased view. The road to general AI for humans is still long and arduous.

Three conditions for artificial general intelligence

AI still has a long way to go, because to achieve artificial general intelligence (AGI), three conditions must be met. The first condition is that the output of AGI must be domain-independent. Although ChatGPT has shown domain-independent characteristics in dialogue and natural language processing, it has not yet fully realized this characteristic in many other issues. The currently developed medical diagnosis system can only diagnose specific diseases and has not explored a general system that can cover the diagnostic functions of all diseases. For the domain-independent ChatGPT technology, it has only taken a step. The second condition is that the system must be task-independent and have the potential to perform various tasks. ChatGPT currently has dialogue capabilities, can also perform arithmetic operations, and can even write poems and code, showing multi-task processing capabilities. However, this does not mean that it can handle all tasks, especially when facing complex tasks in complex environments, it still seems powerless. The third condition is to establish a unified theory.

Four steps from LLM to artificial general intelligence

To move from an LLM towards AGI, four steps need to be completed first (Figure 11). The first step is to interact with humans and align with humans, the second

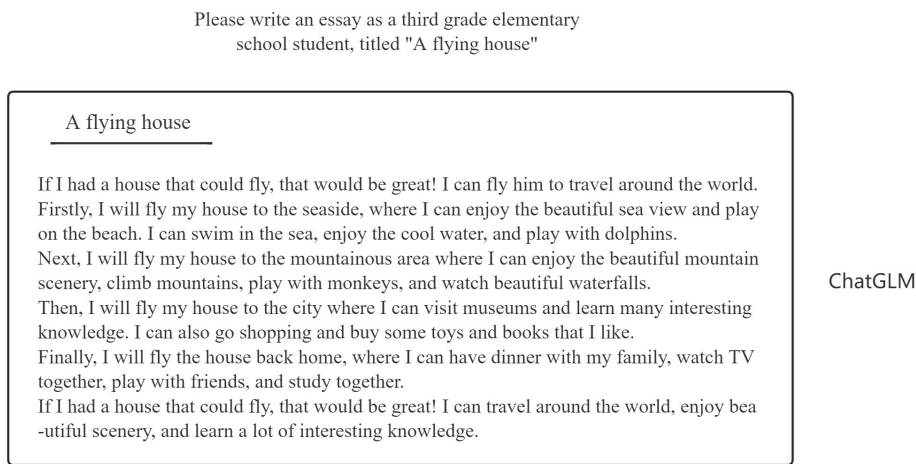


Figure 10. “A flying house”.

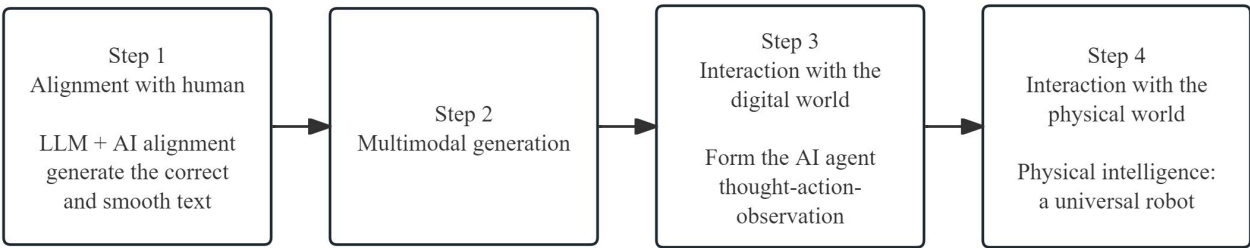


Figure 11. From the LLM to AGI. AI: artificial intelligence; LLM: large language model.

step is multimodal generation, the third step is to interact with the digital world, and the fourth step is to interact with the physical world. Completing these four steps does not mean that artificial general intelligence has been achieved, but they are undoubtedly important steps towards artificial general intelligence.

Step one: align with humans

Although ChatGPT is currently able to output language that humans can understand, this does not mean that its responses are always accurate. To solve this problem, it is necessary to use human wisdom and experience to assist it in making improvements, ensuring that its output is consistent with human understanding and expectations.

For example, the data in Figure 12 shows that at one point, GPT-3 had an error rate of 40% (i.e., 40% errors, 60% correct). After two years of human adjustments to help it correct, the error rate was reduced to 20% during the ChatGPT period, and further reduced to 10% during

the GDP-4 period. These data clearly show that the errors in ChatGPT do indeed need human assistance to correct, and the speed of this correction is quite fast. Although the rate of iteration is fast, it is also important to recognize that errors cannot be completely eliminated. In order to endow ChatGPT with creativity, it is necessary to accept the reality that it may make mistakes.

Step two: multimodal generation

With technological advancements, large models can now be applied to generate various modalities, including images, sounds, videos, and code. The range of sound generation covers multiple fields such as speech and music, so large models can generate diversified modal content.

As shown in Figure 13, the image is generated from text by Shengshu Technology, that is, it automatically generates relevant images based on the input text. The picture in the middle shows the theme of the Mid-Autumn Festival's moon rabbit and mooncakes, depicting images of the moon rabbit and mooncakes.

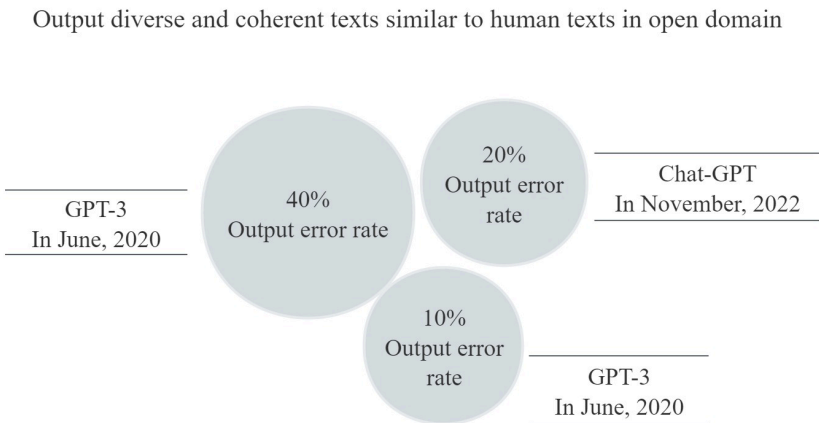


Figure 12. The effect of the AI alignment. AI: artificial intelligence, GPT, Generative Pre-trained Transformer.



Figure 13. Text to Image-Shengshu Tech.

Looking at the rabbit part, its whiskers and details are excellently presented. Of course, images can also be drawn in the specific style of a certain painter. The picture on the left is a sunflower presented in the style of the painter Cézanne, exhibiting excellent quality and reaching a high level of artistic standard.

Currently, some image generators can vividly depict facial expressions, clothing, and background atmospheres in their pictures. On one hand, this reflects their level of artistry; on the other hand, it creates significant potential for fraud. If a work depicts a scene that is alleged to involve misconduct and claims it shows you caught in the act and making a mistake, how can you defend yourself? For this reason, companies are now focusing on verifying the authenticity of text, determining whether it is machine-generated or human-written, and whether videos are made by humans or machines.

As technology progresses, this will become increasingly difficult, providing an excellent opportunity for fraud, a specialty known as "deepfake". It means using deep learning methods to fake, not that the fakes are deep. If in the future 95% of the text content on the Internet is generated by machines, whether people can get the truth from it is still worth thinking about. Once an event sparks public controversy and a large amount of opposition emerges on the Internet, whether this is the majority of people expressing their opinions honestly, or the result of a few people controlling machines to orchestrate, such a possibility will provide convenience for exploiting and manipulating public opinion, therefore, the seriousness of this issue cannot be ignored.

Figure 14 is a 3D image made by our research team, completely generated by a computer. As for the video, it is formed by continuously generating a series of images

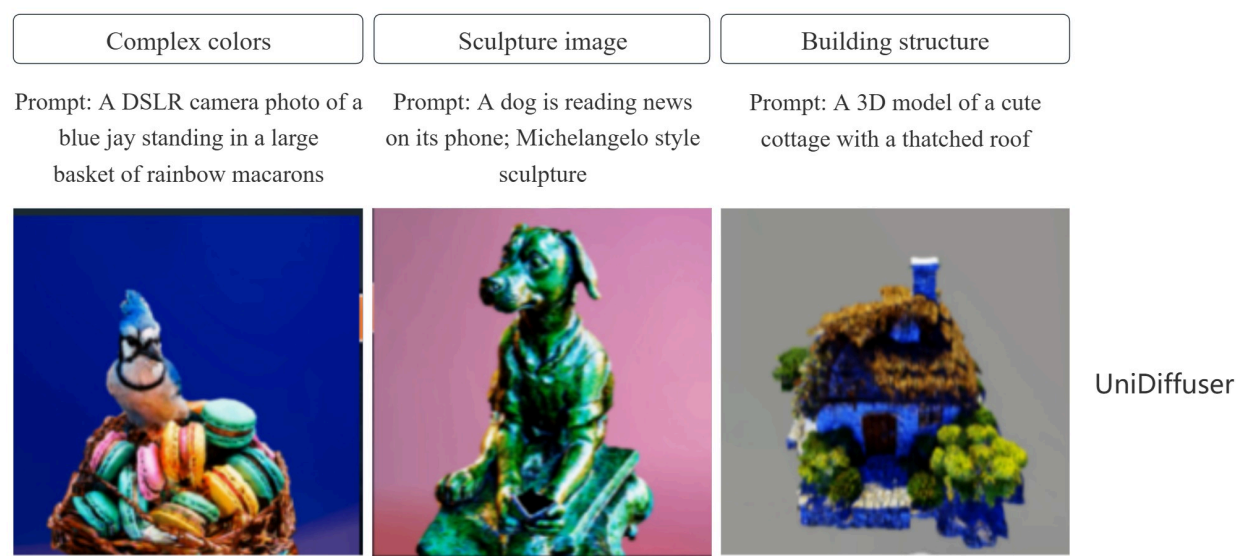


Figure 14. Generating 3D image-Shengshu Tech.

with temporal correlation.

At present, AI has achieved three breakthroughs, one of which is to generate semantically coherent human-like text in open domains. Semantic coherence is the most important breakthrough. With this breakthrough, there is a breakthrough in images, as images only require spatial coherence, while videos require coherence in both time and space. The key lies in the breakthrough at the language level. Based on this, further breakthroughs in image technology can be achieved, and with the breakthrough in image technology, the development of video technology will also usher in new breakthroughs. Of course, the demand for computational resources by images and videos is increasing, and the corresponding computational hardware is also increasing.

The so-called "emergence" phenomenon does not appear until the system scale reaches a certain threshold. In other words, when the scale is small, the presented image is very bad, the horse does not have a horse head, and the painting is very unlike. Only when the system scale reaches a certain level, the image presents a good situation. This is called emergence, that is, from quantitative change to qualitative change. So far, there is still a certain gap in the complete understanding of this phenomenon worldwide, which is also an important reason for causing panic. Many people exaggerate this and worry about whether the computer will have consciousness as the system scale expands, whether it will actively attack humans, or even whether it will come out to rule humans. These concerns stem largely from the difficulty in explaining the "emergence"

phenomenon. However, people don't need to panic about this. Although the "emergence" phenomenon has a certain complexity, it still has some weaknesses.

Step three: AI agents

They must be connected to the digital world. Although LLMs have excellent capabilities, it is far from enough to stay at the level of "speaking". Practical operation is the key. The primary task is to carry out specific tasks in the digital world, to solve actual problems through practical operations, and to understand how they perform in these tasks. This feedback mechanism has significant benefits for improving their performance. In the LLM, computers might only describe verbally and were not sure whether they were correct. But now, AI agents can execute according to human instructions, and they can immediately know whether they are correct once they execute, so obtaining feedback helps to promote the continuous development of large models (Figure 15).

Step four: embodied intelligence

At this stage, it must be related to the physical world. Tasks cannot be completed only with verbal expression and lack of actual action. To interact with the physical world, robots need to be used. This is the concept of "embodied intelligence", that is, intelligence is presented in the form of having a body. Intelligence is not enough just by the brain, it also needs to have a physical body to achieve coordination between verbal expression and action (Figure 16).

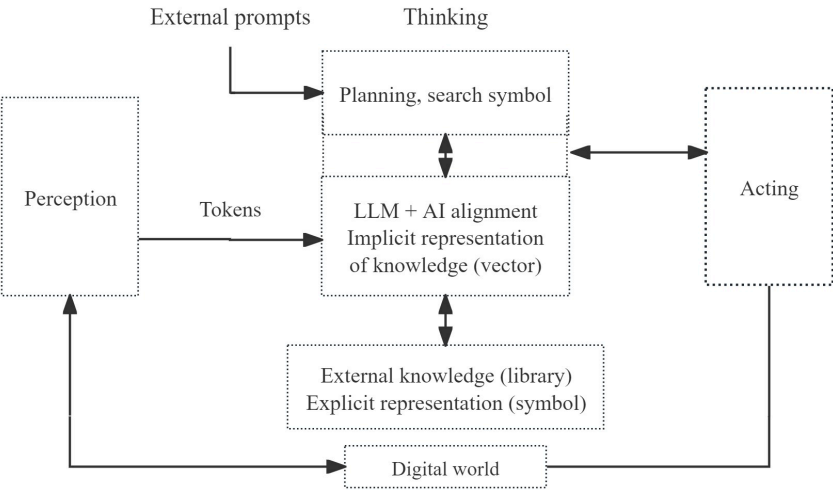


Figure 15. Step three: The AI Agent. AI: artificial intelligence; LLM: large language model.

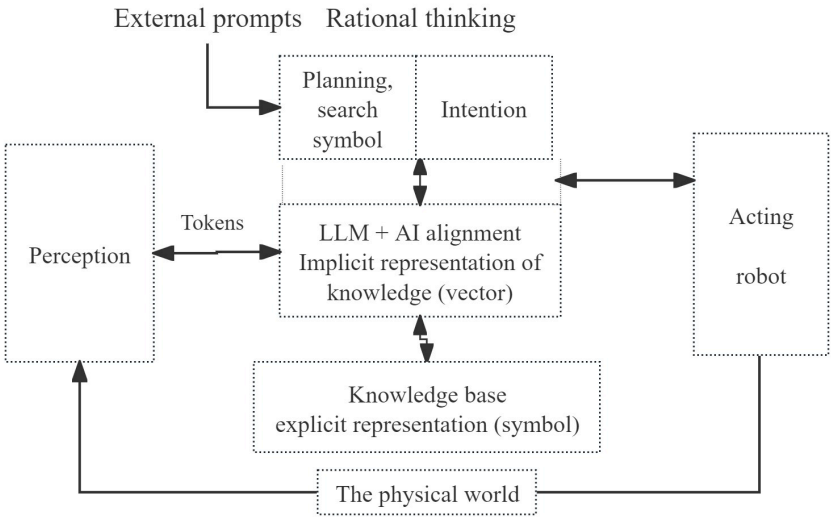


Figure 16. Step four: Embodied intelligence. AI: artificial intelligence; LLM: large language model.

CONCLUSION

The industrial development of AI

The rapid development of the information industry benefits from its solid theoretical foundation. Based on this theory, the developed hardware and software are universal. Universality means a huge market space, so large companies such as Intel, IBM, and Microsoft have emerged one after another, promoting the application and promotion of technology, promoting the development of informatization, and the entire industry chain shows a rapid growth trend. In contrast, the

development of the field of AI is somewhat difficult. It lacks the systematic theoretical support like the information industry. It currently mainly relies on the construction of algorithms and models. The hardware and software built according to these algorithms and models are all special. Special means relatively small market demand. So far, there has been no giant like IBM, Intel, and Microsoft in AI. Therefore, the development of the AI industry must be deeply integrated with the vertical field to promote its development. However, the current situation has changed, and basic models with a certain degree of

universality have begun to emerge, which may have an important impact on the development of the industry.

Regarding the current development of the AI industry, according to statistics, there were 40 unicorn companies with a valuation of more than 1 billion United States dollars worldwide in 2020. By 2022, this number had significantly increased to 117, By early 2024, the number of such companies had climbed to 126. It can be seen from this that the AI industry is showing a steady growth trend.^[5]

Regarding the future development trend of foundation models, currently, China's foundation model scene is described as a "battle of hundreds of models", with 100 entities or even more, while in the United States, it is basically a few entities (such as Meta, Google, and OpenAI). Although our country has many entities engaged in the construction of general foundation models, they will gradually face difficulties. There are three possible directions for them (Figure 17). The first direction is to shift to various industries and develop foundation models for various vertical fields. Many industries are considering this issue now. The oil industry will inevitably focus on foundation models in the oil field, and the financial industry will focus on foundation models in the financial field. In the future, the number of companies focusing on building general-purpose foundation models will gradually decrease, and more companies may shift their focus to the development of foundation models in various vertical fields. The second direction is to apply the models to the industry after fine-tuning. They can provide open foundation model software for people to develop applications. The third direction is to combine with other technologies to develop new industries. For example, text editing was done with software before, but now with the addition of foundation models, it can help with drafting. Many journalists use this tool, first letting it assist in completing the first draft of the manuscript, and then doing the subsequent editing work, which undoubtedly greatly improves work efficiency. The integration of other technologies with existing industries can lead to the development of new industries.

Taking China as an example, the foundation model developed by the Zhipu AI company (incubated by Tsinghua University) has shown a good development momentum, and the development of other foundation models in China has gradually received attention and recognition. This will undoubtedly drive profound changes in the industry. In the future, whether engaged in hardware or software development, this platform should be used. In the past, software development was often carried out in a computer environment without any foundation, which is like teaching a person who

knows nothing to complete complex tasks, and the efficiency is naturally low. Now, if this work is put on the platform of the foundation model, the situation will be very different. This platform has already learned from 13.51 million Oxford dictionaries, and its knowledge reserve is at least equivalent to the level of a high school student. Therefore, the use of such a platform has become an unstoppable trend. These "high school students" are provided by foundation model companies' public platforms for everyone to use.

The limitations of foundation models

Whether foundation models are universal, what are the restrictions and main challenges for foundation models (Figure 18)? All tasks of foundation models are driven by external factors, that is, they are executed after receiving external prompts. First of all, foundation models lack initiative and need to act under external prompts. When accepting external guidance, they use probabilistic prediction methods to perform tasks, which may bring some defects that humans do not have, that is, the quality of output is uncontrollable, and it does not have the ability to judge whether its output is correct or not, so its output results are often difficult to trust. Secondly, it is significantly affected by the external environment, and its behavior is usually directly guided by external requirements, while humans show obvious self-consciousness, even if the task is assigned by others, humans can still complete it under the guidance of self-consciousness, making its performance has controllability and reliability. In contrast, machines lack cognition of their own activities and cannot be aware of their behavioral intentions. This can be illustrated by the following two cases.

Case one

when asked about the lyrics of Tsinghua University's school song, ChatGPT doesn't know and even made up a set of lyrics out of thin air. Point out its mistake: "The school song of Tsinghua University is not this, it is 'West mountains are vast, the East Sea is boundless...'" Copy these two sentences to it, followed by an ellipsis, hoping that it can be fully supplemented. It said: "Sorry, I was wrong just now, the school song of Tsinghua University is 'West mountains are vast, the East Sea is boundless'." As a result, it went on to make up the following. Point out its mistake again and tell it: "You are wrong, the school song of Tsinghua University is not this." Copy the entire original text to it, and it immediately says: "Sorry, everything I said was wrong, it should be this". However, whenever you exit the conversation and ask again, it will make up the lyrics again. It cannot judge right from wrong by itself, and requires human intervention and correction in the background. This rules out a problem, some people worry that it will gradually iterate and get better and better, but in fact,

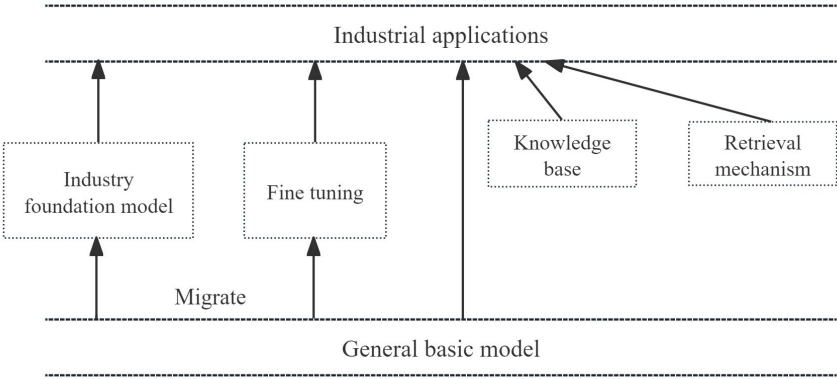


Figure 17. Different application modes of the basic model.

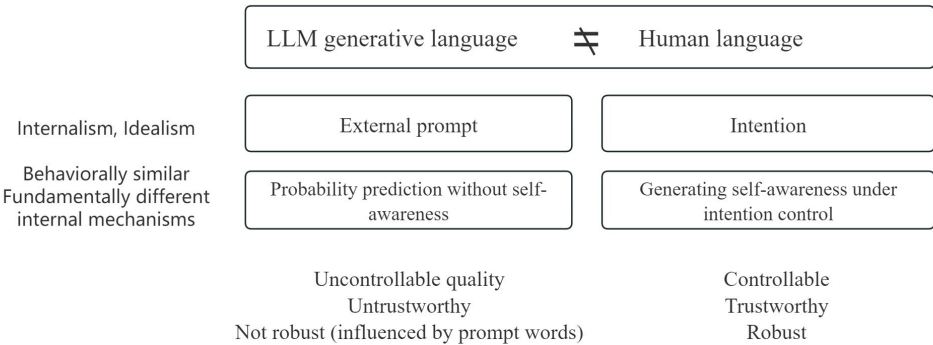


Figure 18. Limitations of the foundation model. LLM: large language model.

because it cannot modify and improve itself, and cannot judge right from wrong. As for whether it will take the initiative in the future, this issue is still under study, but what can be confirmed at present is that it cannot do this yet. Whether it has self-motivation to act and whether it will generate such a force are still under study, and there will be no definite answer in the short term. So iteration is done with the help of humans, and it currently does not have the ability to self-iterate.

Case two

Its output is not always ideal (Figure 19). Works like Sora are often the high-quality results that are shown after selection. Take the six images generated by the famous American image generator DALL·E.2. As an example, after inputting "He angrily kicked the door open and walked out in a lofty manner", the first three artworks are satisfactory, but the subsequent works not only do not meet the theme requirements, but also the quality of creation is poor. It will produce poor quality work, which is also an important reason for people to have doubts about its credibility. In the future, it will

mostly assist people in completing work, or part of the work is done by machines with human monitoring. At present, there is only a small amount of work that can be done entirely by machines.

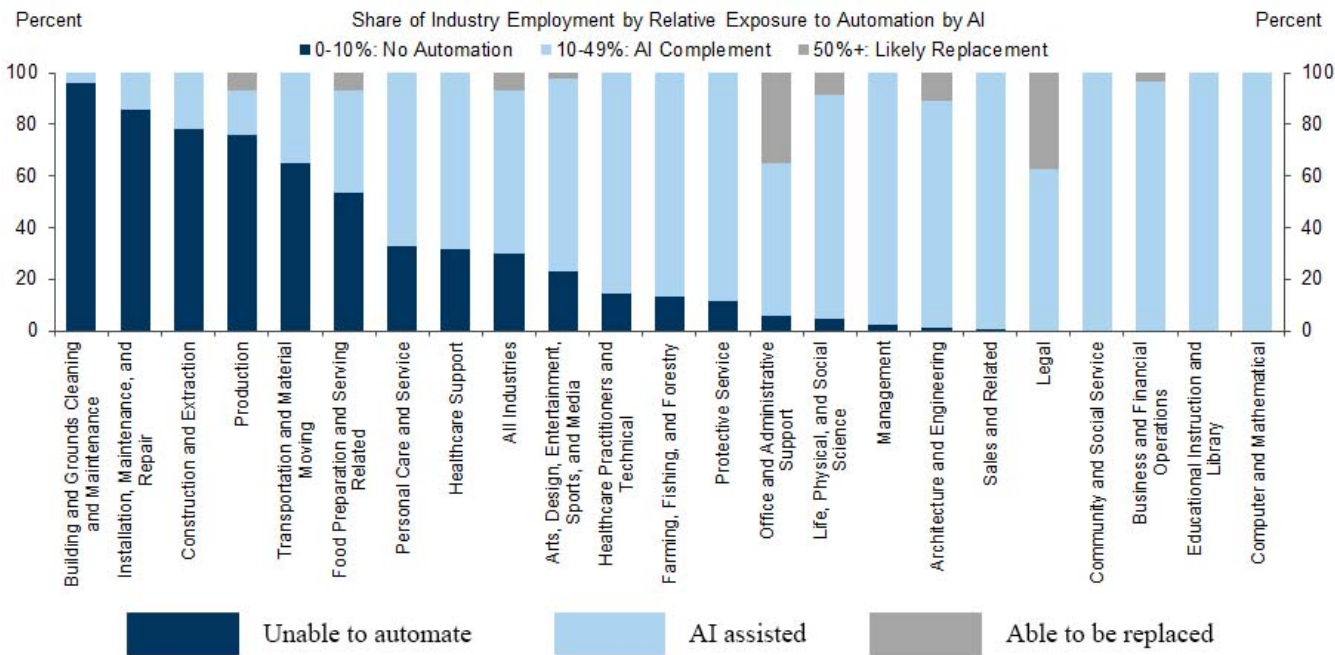
Goldman Sachs Global Investment Research has created a statistical chart that provides a detailed analysis of the impact of AI on various industries. Figure 20 lists many industries and uses shades of color to reflect the extent of AI's impact on them. Deeper shades of blue indicate a smaller impact, while lighter shades of blue indicate a larger impact. The grey areas in the chart represent those industry sectors that may be replaced by AI. It is clear from the chart that the proportion of grey areas is relatively small, meaning that only a minority of jobs will be replaced. Jobs that may be replaced in the future include administrative work, secretarial work, and legal work involving a large amount of document sorting. While AI has had a significant impact on all industries, in most cases, it has helped people improve the quality and efficiency of their work. Only a few jobs will be completely replaced.



DALLE.2

"He kicked off to open the door angrily and walked away"

Figure 19. Output quality is inconsistent.



Source: Goldman Sachs Global Investment Research, 2023.3

Figure 20. The impact of AI on various industries. Adapted from Hatzius *et al.*^[6] AI: artificial intelligence.

The three-space model of the third generation of AI

How to develop AI theory is an important issue currently faced by humanity. In the past, text was processed in one space, and images and voices were processed in another space, with no effective connection between the two spaces. Now, with the emergence of foundation models, a middle space has been formed, connecting the whole set of perception and cognition,

providing excellent conditions for the development of AI theory (Figure 21).

AI is like an unexplored "no-man's land", its charm lies in the fact that it is always on a journey of continuous exploration and advancement. For the development of AI, we should not be blindly optimistic due to temporary progress, nor should we be disheartened by setbacks. On the contrary, we should maintain enduring

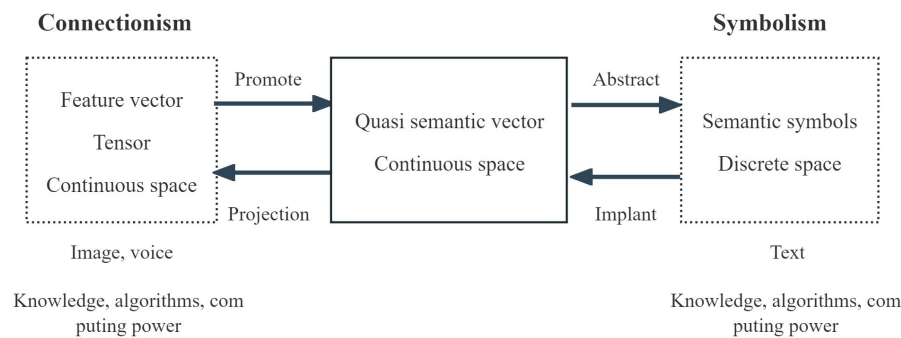


Figure 21. A three-space model of the third generation of AI. AI: artificial intelligence.

enthusiasm and perseverance, and continuously put in effort.

DECLARATIONS

Author contributions

Zhang B: Conceptualization, Writing—Original draft preparation, Writing—Reviewing and Editing, Project administration. The author has read and approved the final version of the manuscript.

Source of funding

This research received no external funding.

Ethical approval

Not applicable.

Informed consent

The author declares that he has obtained appropriate informed consent from persons or their guardians appeared in the figures to be published in this article.

AI usage statement

Declaration: To make this paper more vivid and illustrative, during the preparation of this paper, Figure 10 was generated using AI-ChatGLM, Figure 13 was generated using AI-Image-Shengshu Tech, Figure 14 was generated using AI-Generating 3D image-Shengshu

Tech, and Figure 19 was generated using AI-DALLE.2. I have reviewed and edited the content as needed.

Conflict of interest

The author has no conflicts of interest to declare.

Data availability statement

No additional data.

REFERENCES

1. Koch A, Cascorbi I, Westhofen M, Dafotakis M, Klapa S, Kuhtz-Buschbeck JP. The neurophysiology and treatment of motion sickness. *Dtsch Arztebl Int.* 2018;115(41):687-696.
2. Xinhua News Agency. [Government work report]. The Central People's Government of the People's Republic of China. Updated March 5, 2024. Accessed June 5, 2024. https://www.gov.cn/yaowen/liebiao/202403/content_6939153.htm
3. Dong Y, Liao F, Pang T, et al. Boosting Adversarial Attacks with Momentum. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE; 2017: 9185-9193.
4. CSRankings: Computer Science Rankings. Emery Berger. Accessed May 21, 2024. <https://csrankings.org/#/fromyear/2017/toyear/2019/index?all&cn>
5. Xinhua News Agency App. [Global AI welcomes a new round of rapid development]. Baidu. Updated August 11, 2023. Accessed May 21, 2024. <https://baijiahao.baidu.com/s?id=1773893896996879444&wfr=spider&for=pc>
6. Hatzius J, Briggs J, Kodnani D, Pierdomenico G. *The potentially large effects of artificial intelligence on economic growth*. Goldman Sachs; 2023.