AI Overview

HUAWEI

Foreword

- Artificial intelligence (AI) is driving progress and industrial upgrades at an unprecedented speed. As a field of research in human intelligence simulation, AI has not only profoundly changed our lifestyles but has also showcased its immense value across various fields such as research, industrial production, healthcare, financial services, and education and entertainment.
- This section describes the history, key technologies, and applications of AI in various fields. It provides insights into the application fields and future of AI technologies based on their definition and scope.

Objectives

- Upon completing this course, you will be able to understand:
  - Basic AI concepts
  - AI technologies and their history
  - AI applications in various technologies and fields
  - Future AI trends

Contents

AI in the Eyes of Researchers

"I propose to consider the question, 'Can machines think?'"

-- Alan Turing in 1950

Make machines behave like humans.

-- John McCarthy in 1956

"Artificial intelligence is the science of making machines do things that would require intelligence if done by men."

-- Marvin Minsky

HUAWEI

- "The branch of computer science concerned with making computers behave like humans." -- A widespread early definition of AI proposed by John McCarthy at the Dartmouth Conference in 1956. However, it seems that this definition ignores the possibility of strong AI. According to another definition, AI is the intelligence (weak AI) demonstrated by artificial machines.
- Alan Turing discussed the question "Can machines think" in his paper Computing Machinery and Intelligence.

What Is Intelligence?

- Professor Howard Gardner proposed the theory of multi-intelligence, and listed eight capabilities that reflect multi-intelligence:
  - Verbal/Linguistic
  - Logical/Mathematical
  - Visual/Spatial
  - Bodily/Kinesthetic
  - Musical/Rhythmic
  - Inter-personal/Social
  - Intra-personal/Introspective
  - Naturalist

HUAWEI

- Gardner, American educator and psychologist
- 1. Linguistic intelligence

It refers to the ability to express thoughts and understand others orally and literally, to grasp speeches, semantics, and grammars masterly, and to think in words, express in words, and appreciate the deep meaning of languages. They are competent for jobs such as political activists, presenters, lawyers, speakers, editors, writers, journalists, and teachers.

- 2. Logical-mathematical intelligence

It refers to the ability to calculate, measure, infer, conclude, classify, and carry out complex mathematical operations. This type of intelligence possesses sensitivity to logical means and relationships, statements and propositions, functions, and other related abstract concepts. They are competent for jobs such as scientists, accountants, statisticians, engineers, and computer software R&D personnel.

- 3. Spatial intelligence

It refers to the ability to accurately perceive the visual space and surroundings and present the perception in the form of graphics. This type of intelligence possesses sensitivity to colors, lines, shapes, forms, and spatial relationships. They are competent for jobs such as interior designers, architects, photographers, painters, and pilots.

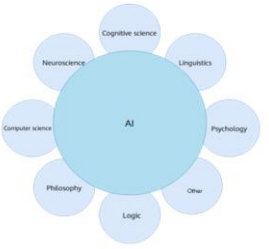
- 4. Bodily-kinesthetic intelligence

It refers to the ability to express thoughts and feelings with the body and to make or operate

objects with hands. This type of intelligence possesses special physical skills such as balance, coordination, agility, strength, flexibility and speed, and abilities triggered by tactile sensation. They are component for jobs such as athletes, actors, dancers, surgeons, jewellers, and mechanics.

What Is AI?

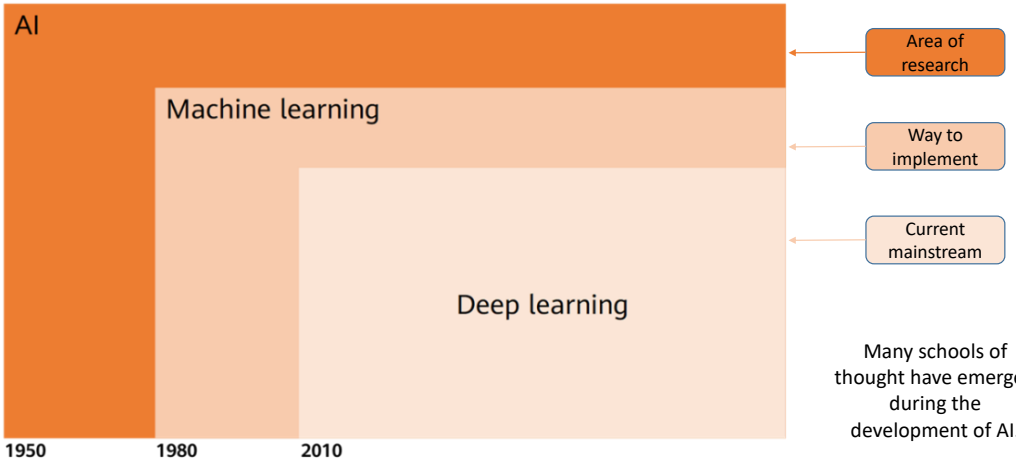
- Artificial intelligence can be interpreted as "artificial" plus "intelligence". "Artificial" means being designed, created, and made by humans. "Intelligence" means thinking and behaving like humans.
- Artificial intelligence is a new technical science that deals with the research and development of theories, methods, techniques, and application systems for simulating and extending human intelligence. In 1956, the term "Artificial intelligence" was first coined by John McCarthy, who defined it as "the science and engineering of making intelligent machines." The purpose of AI is to make machines intelligent and give them human thoughts.
- AI has become an interdisciplinary subject that overlaps with various fields.



HUAWEI

- Machine learning can be understood from multiple aspects. Tom Mitchell, known as the "godfather of global machine learning", defined machine learning as: For a task T and performance metric P, if the performance of a computer program measured by P on T self-improves with experience E, the computer program is learning from experience E. These definitions are simple and abstract. However, as we deepen our understanding of machine learning, we will find that the connotation and extension of machine learning are changing over time. Because a variety of fields and applications are involved and machine learning develops rapidly, it is not easy to define machine learning simply and clearly.

Relationships Among AI, Machine Learning, and Deep Learning (1)



Many schools of thought have emerged during the development of AI.

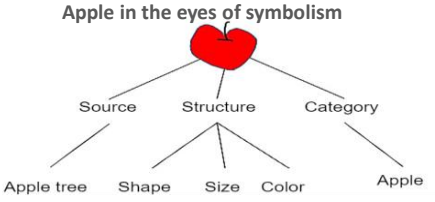
Relationships Among AI, Machine Learning, and Deep Learning (2)

- AI is a new technical science that studies and develops theories, methods, techniques, and application systems to simulate and extend human intelligence.
- Machine learning is the study of how computers simulate or replicate human learning processes to acquire new knowledge or skills and restructure existing knowledge for continuous performance improvement.
- Deep learning is a concept that originates from the research of artificial neural networks. It is an ongoing field of machine learning research that mimics the human brain's mechanism to interpret data.

- The term "artificial intelligence" was previously used to describe machines that imitate and demonstrate "human" cognitive skill related to human thinking.
- A machine learning algorithm mainly builds a model based on sample data (referred to as trained data), so that prediction or decision-making can be made without explicit programming.
- Deep learning is a type of machine learning, and machine learning is a one and only path for implementing artificial intelligence. The concept of deep learning originates from the study of artificial neural networks. A multi-layer perceptron containing multiple hidden layers is a deep learning structure. Deep learning uses higher level features derived from the lower level features to form a hierarchical representation. The motivation of deep learning research is to establish a neural network that simulates the human brain for analysis and learning. The neural network simulates the mechanism of the human brain to interpret data, such as images, sounds, and texts.

Major Schools of AI - Symbolism

- Symbolism is also called Logicism, Psychologism, or Computerism.
- Symbolism believes that AI is built upon mathematical logic. Followers of this school of thought believe that symbols are the cognitive primitives of humans, and the human cognition process is based on inferring and calculating a variety of symbols. In their opinion, both humans and computers use physical symbol systems, so computers can be used to simulate the intelligent behavior of humans.

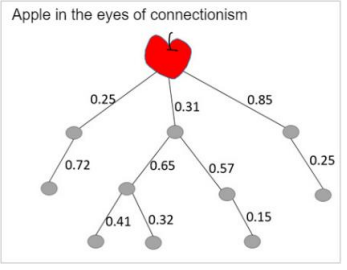


HUAWEI

- After 1956, symbolism developed heuristic algorithms, expert systems, and knowledge engineering theories and technologies, and made significant progress in the 1980s.
- The successful launch and application of the expert system is of great significance for leading AI to engineering application and linking AI theory with practice.
- After other AI schools emerged, symbolism is still a mainstream of AI.

Major Schools of AI - Connectionism

- Connectionism is also called Bionicsism or Physiologism.
- Connectionism believes that AI originates from bionics, particularly the research on human brain models. The primitives of human thinking are neurons, rather than symbolic processes. Researchers in this school of thought began by looking at neurons and then neural network models and brain models. This has successfully opened another path to artificial intelligence.

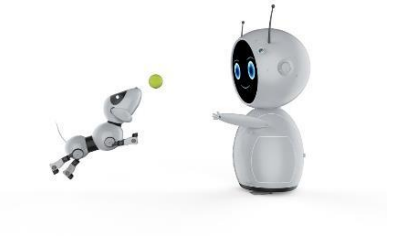


HUAWEI

- Units in a network can represent neurons, and connections can represent synapses, just as in the human brain.
- In 1986, Rumelhart et al. proposed a backpropagation (BP) algorithm in a multi-layer network. Since then, connectionism has gradually emerged, from models to algorithms, and from theoretical analysis to engineering implementation.

Major Schools of AI - Behaviorism

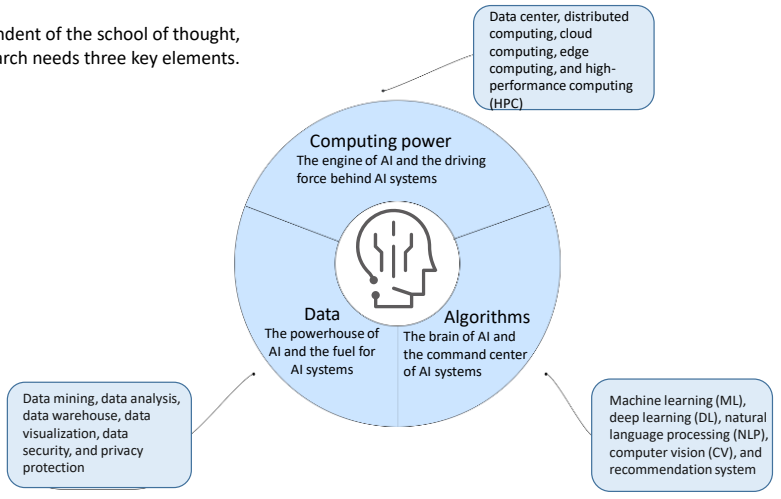
- Behaviorism is also called Evolutionism or Cyberneticsism.
- Behaviorism believes that AI originates from cybernetics. In this school of thought, intelligence depends on perception and action; and knowledge, representation, and inference are not involved. AI can evolve like human intelligence. Intelligent behavior can only be manifested through interaction with surrounding environments in the real world.



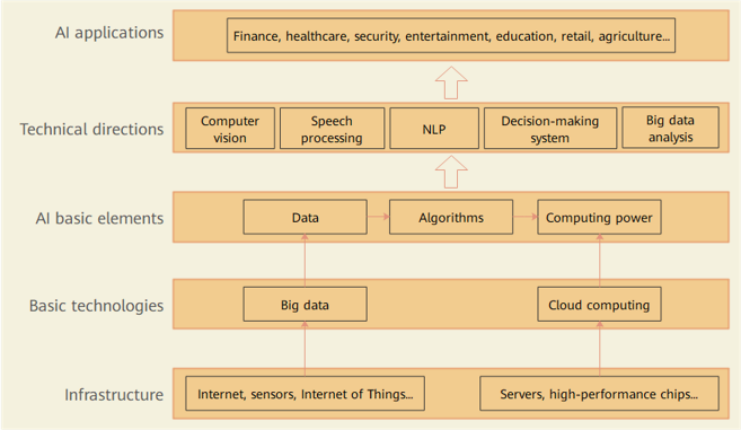
- Behaviorism concerns more about application practices and how to learn from the environment continuously to make corrections.
- School masterpiece: Brooks' hexapod walking robot, a control system that simulates insect behavior based on the perception-action pattern.
- This school is a somewhat similar to an adaptive control system, which collects data using sensors (environments) and acts on the system.

Three Elements for AI Development

Independent of the school of thought, AI research needs three key elements.



AI Industry Ecosystem



HUAWEI

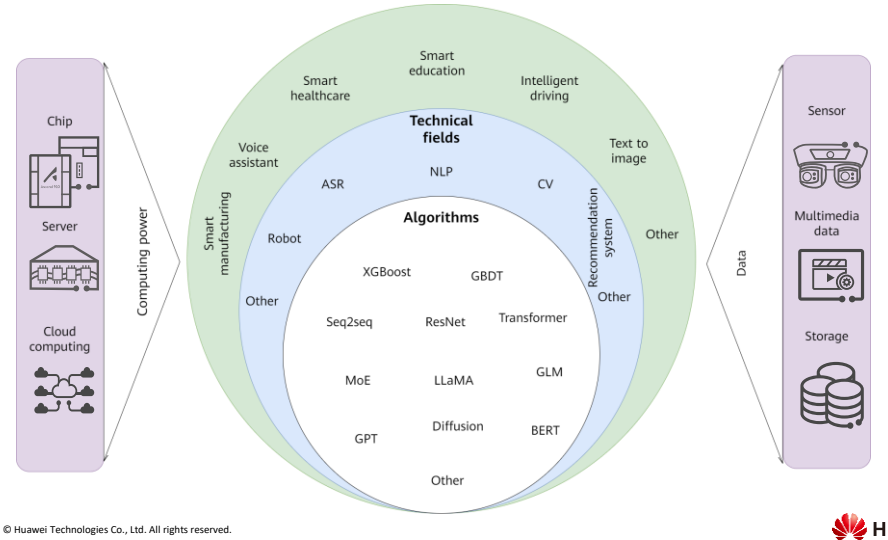
Types of AI

- Strong AI
  - The strong AI theory is that it is possible to create intelligent machines capable of reasoning and problem solving, and that such machines will be self-aware. They will be able to think independently and develop the best solution to a problem. In addition, such machines will have their own values and worldview, as well as the instincts of living things, such as the need for survival and safety. In a sense, the machine with human thoughts can be regarded as a new civilization.
- Weak AI
  - The weak AI theory holds that it is impossible to create an intelligent machine capable of reasoning and problem solving. Such a machine only looks intelligent but does not have human intelligence or self-awareness.

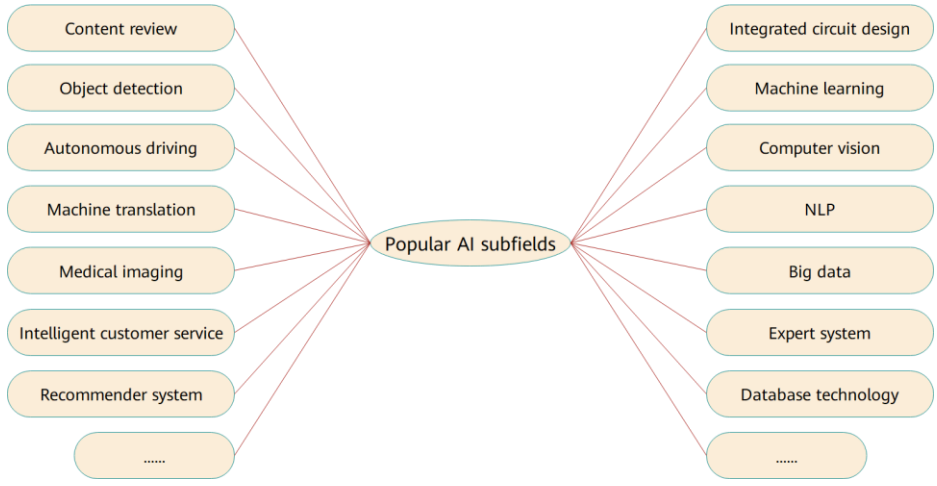
- An important indicator of AI is to achieve a superhuman level in challenging fields through self-learning without any prior knowledge.
- Strong AI refers to AI that can compete with humans in all aspects. Therefore, strong AI is not limited to a specified field, but makes robots comparable to humans in all aspects. Strong AI can think, plan, solve problems, abstract thinking, understand complex concepts, quickly learn, and learn from experience. Currently, it is believed that if the human brain can be simulated and all neurons and synapses in the human brain can be imitated on the same scale, strong AI will naturally occur.
- Now we are in the weak AI phase. Weak AI alleviates human intellectual labor, similar to advanced bionics. AI outperforms humans only in some aspects.

Contents

Overview of AI Technologies & Application Fields



Popular AI Subfields



Distribution of AI Applications

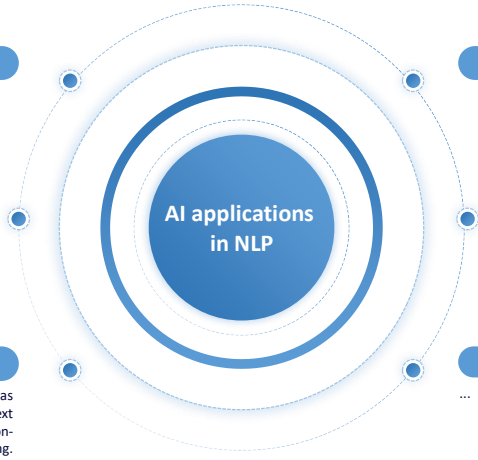
- AI is commonly applied in the following technical fields:
  - Computer vision
    - How can we enable computers to "see" and "see" faster and more accurately?
  - Natural language processing
    - How can we enable computers to understand and use natural languages?
  - Multimodal
  - Other

Contents

Natural Language Processing

- Natural language processing (NLP) is a branch of computer science and artificial intelligence. It studies theories and methods for implementing effective communication between humans and computers through natural languages. NLP is intended to enable computers to understand and generate human languages.
  - Development: (1) Rules and linguistic theories; (2) Machine learning and deep learning; (3) Large language models (LLMs)
- Applications:
  - Sentiment analysis
  - Text mining
  - Machine translation
  - Chatbots
  - Other

AI Applications in NLP



Text classification

Automatically classifies text into preset categories by using algorithms, and analyzes text features to manage information efficiently and support precise decision-making.

Text summarization

Automatically extracts core information from text and generates a more concise version while retaining the original semantic associations and key elements.

Information extraction

Automatically extracts key information, such as entities and relationships, from massive text data to support data analytics and decision-making.

Machine translation

Automatically converts a language into another language (for example, Chinese to English, or French to Japanese) using algorithm models.

Smart writing

Assists in generating high-quality text, such as news releases and advertising copies, improving content creation efficiency.

...

...

NLP Task - Text Classification

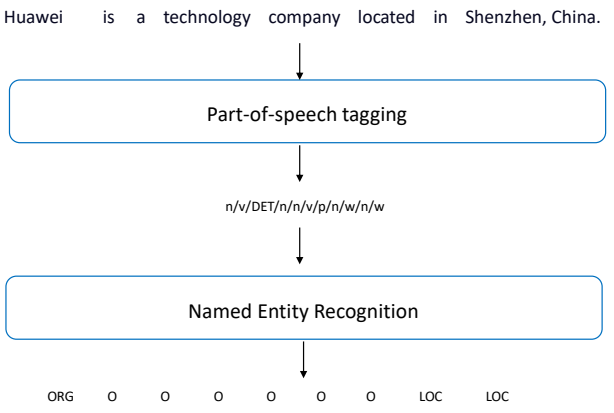
- Definition
  - Input:
    - An article d
    - Fixed category set $C = \{C_1, C_2, \dots, C_j\}$
  - Output:
    - Predicted category for the article $c \in C$

I wasn't happy to hear about a price cut a few days after this purchase. The flash memory score is just over 500 points. ☹

Pleasing appearance and amazing sound. Now, most electronic products are made by Huawei. 😀

NLP Task - Sequence Labeling

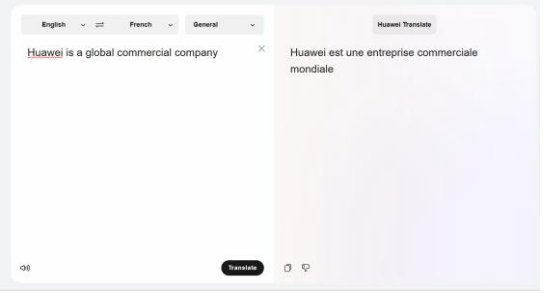
- Sequence labeling is using a model to label each position of a given input sequence (usually a sentence, character, or word in text) to form a label sequence. This process may be viewed as sequence-to-sequence mapping.



- DET represents determiner, n represents noun, v represents verb, p represents preposition, and w represents punctuation.

NLP Task - Text Generation

- Text generation is a process in which an algorithm automatically generates grammatically sound and meaningful text content.
- This technology can simulate the human writing process and generate various types of text, such as news reports, emails, articles, social media posts, poems, novel chapters, and machine translations.



Knowledge Graph

- A knowledge graph is a graphical data model used to represent and organize structured information. It represents entities (such as people, places, and things) in the real world and their relationships (such as friendships and affiliations) in the form of graphs to construct a large-scale, multi-domain knowledge base.
- NLP provides the means to build a knowledge graph. NLP helps extract information (such as entities, relationships, and attributes) from text to build and enrich knowledge graphs.

Contents

Computer Vision

- Computer Vision (CV) is a technology that uses computers and algorithms to simulate the process of human vision. It involves extracting information from images or videos, analyzing and comprehending content, and making decisions accordingly.




AI Applications in CV

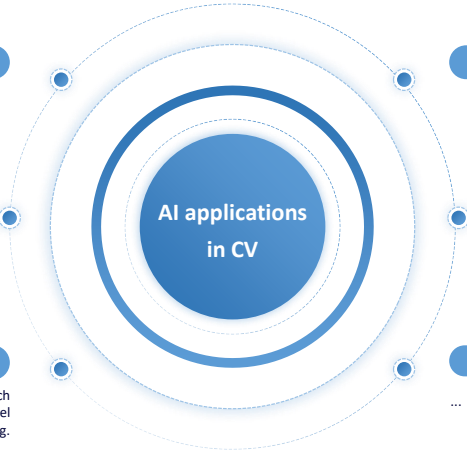


Image classification

Uses extensive amounts of unlabeled data to improve the representation capability of models for applications in various general scenarios.

Object detection

Identifies and locates objects in images or videos, supporting automatic analysis and decision-making.

Semantic segmentation

Assigns a predefined class label to each pixel in an image, implementing pixel-level image understanding.

Video inpainting

Restores and enhances video quality using deep learning-based techniques.

Intelligent driving

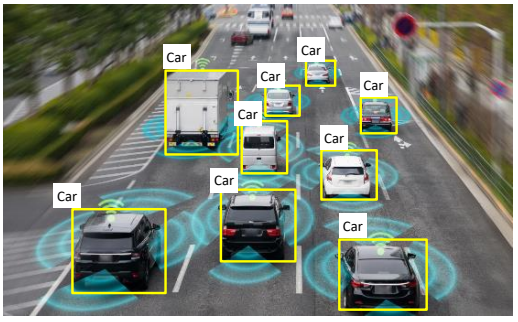
Enables vehicles to perceive surroundings, make decisions, and perform driving operations like human drivers.

...

...

CV Task - Object Detection

- Object detection is one of the fundamental technologies for image processing and computer vision in the AI field. It has a wide range of applications, such as traffic monitoring, image search, facial recognition, and human-computer interaction. Object detection technologies can help detect objects in an image for further processing with intelligent algorithms.

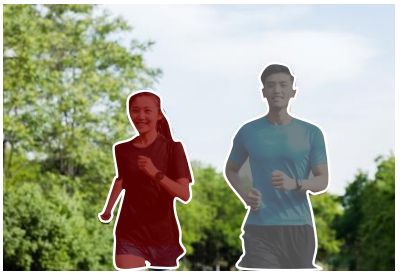


HUAWEI

- Due to different appearances, shapes, and arrangements of objects, as well as lighting and shading during imaging, object detection has always been the largest challenge in the field of computer vision.
- Despite similar techniques used, object recognition is slightly different from object detection. Given a specific object, target recognition is aimed to find an instance of the object in an image. This is not to classify but to judge whether the object appears in the image, and if so, the object is located. For example, a real-time image taken by a security camera is monitored to identify a face of a person.

CV Task - Image Segmentation

- Image segmentation is the process of partitioning an image into multiple segments based on the problem to be solved.
- There are many algorithms and application methods for image segmentation. Common ones include connected component segmentation, motion segmentation, and object segmentation.



HUAWEI

- Object segmentation: semantic segmentation and instance segmentation.
  - Semantic segmentation refers to pixel-level image recognition, that is, marking the object category to which each pixel in the image belongs.
  - Instance segmentation is a task of identifying each instance of an object from an image and performing **pixel-level** labeling on each of the instances.
- Segmentation is a pixel-level description of an image, and makes each pixel category (instance) count. It is applicable to scenarios with high requirements for understanding, for example, segmentation between roads and off-roads in unmanned driving.
- Input and output:
  - Input: image.
  - Output: segmented images with the same resolution as the input image and label of each pixel category.

CV Task - Object Tracking

- Object tracking is a core research area in computer vision and has a wide range of applications, such as intelligent transportation, security monitoring, human-machine interaction, and autonomous driving.
- Tracking algorithms can obtain the trajectories of target objects in terms of time to analyze their movement behaviors.

CV Task - OCR

- Optical Character Recognition (OCR) is the process of recognizing characters in images or scans and converting them into editable ones using image processing technologies such as character object detection and object classification. OCR improves service efficiency by sparing manual information input. OCR can be used to identify characters in ID cards, driving licenses, vehicle licenses, invoices, customs forms, general tables, and general text.

CV Task - Image Generation

- Image generation refers to generating a wholly new image, modifying a part of an image, or generating a new image based on an existing one.
  - Super resolution is a process of estimating a high-resolution image from a low-resolution counterpart and predicting image features at different magnifications.
  - Style migration is a process of applying the style of one domain or several images to other domains or images. For example, an abstract style is applied to a realistic image.
  - Image inpainting is the process of fixing images, for example, restoring damaged black-and-white photos and videos.

Image inpainting

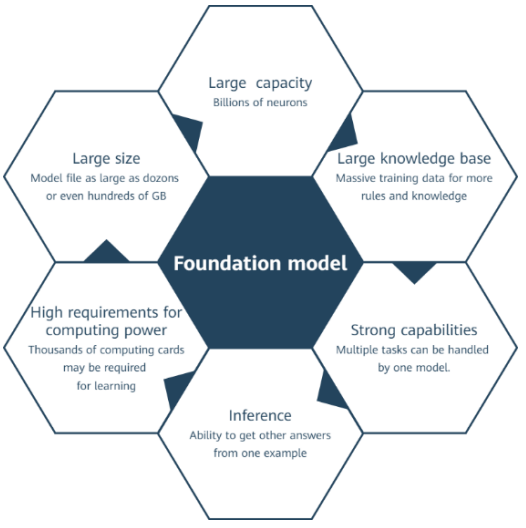
HUAWEI

- Input and output:
  - Input: realistic image and random noise
  - Output: generated image

Contents

What Is a Foundation Model (1)?

- A foundation model refers to a model that is trained on large-scale data, has a massive number of parameters, and has powerful functionalities.



What Is a Foundation Model (2)?

Principles Behind Large Models

Scaling Law

As the model size increases exponentially, the model performance increases linearly.

Chinchilla Law

The model size and the number of training tokens should scale at equal rates.

Emergent Abilities

When the model reaches a certain size and scale, it shows a sudden and unexpected improvement in performance and generalization capability.

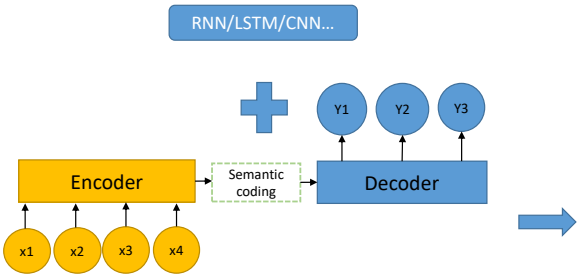
What makes a model "large":

- ✓ Large-scale data: Extensive data to cover nearly all real-world scenarios
- ✓ Large-scale model: High capacity to adapt to almost any scenario
- ✓ Large-scale compute: Robust computing resources for handling intricate calculations

Three problems of large models:

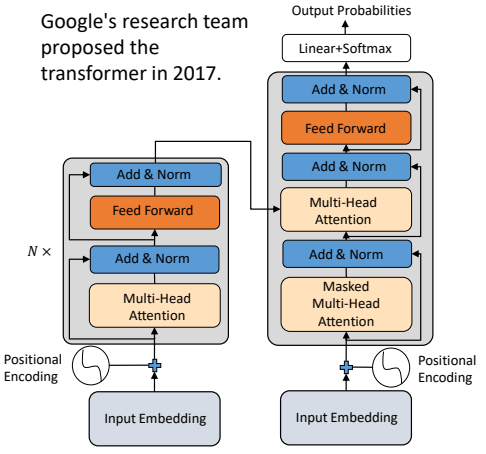
- ✓ How many parameters does a large model have?
- ✓ How can we develop complex algorithms?
- ✓ How does a model process the input data and produce the output?

Where Does a Foundation Model Come from? (1)



In 2014, researchers in machine translation proposed the Seq2Seq model. This was a new way to implement end-to-end machine translation based on the RNN architecture.

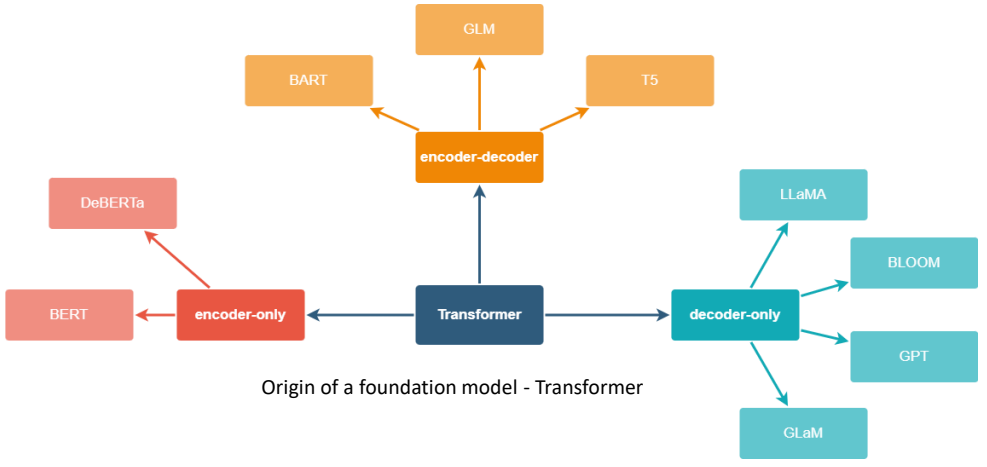
Google's research team proposed the transformer in 2017.



HUAWEI

- The core idea of the Seq2Seq model is to use an encoder network to encode an input sequence into a vector in a fixed dimension or a series of hidden states, and then the decoder network generates a target sequence word by word.

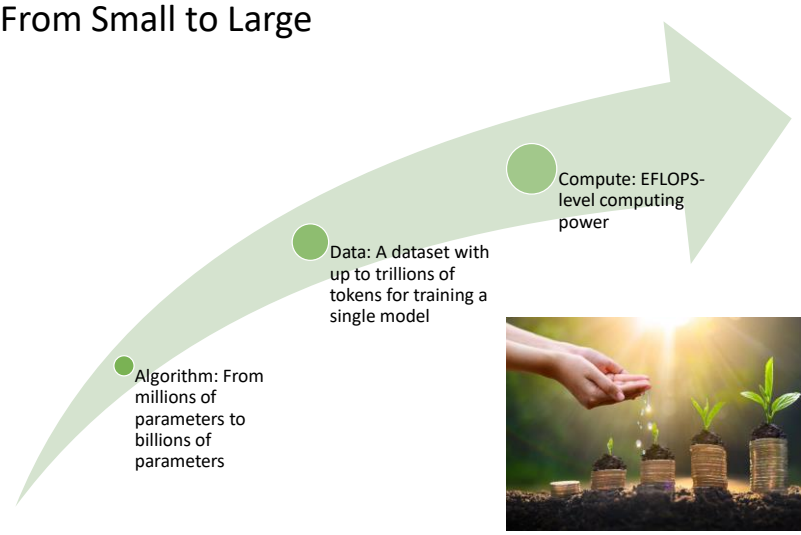
Where Does a Foundation Model Come from? (2)

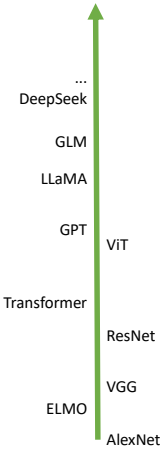


Origin of a foundation model - Transformer

- With Transformer as the infrastructure, foundation models are developed in different branches.

From Small to Large





HUAWEI

Sizes of Foundation Models

- A foundation model usually has hundreds of millions of parameters and has powerful representation and generalization capabilities.

Model	Parameters (million)
ResNet101	44.55
ResNet152	60.19
Yolo v7	74.8
VGG16	138

Model	Parameters (B)	tokens
GPT3	175	3000B
GPT4	17,60 (estimated)	13T
LLaMA1	7–70	1.4T
LLaMA2		2T
LLaMA3	8–70	15T
GLM	6–130	400B
T5	11	34B
PaLM2	340	3.6T
Stable LM 2	21	2T

Dataset

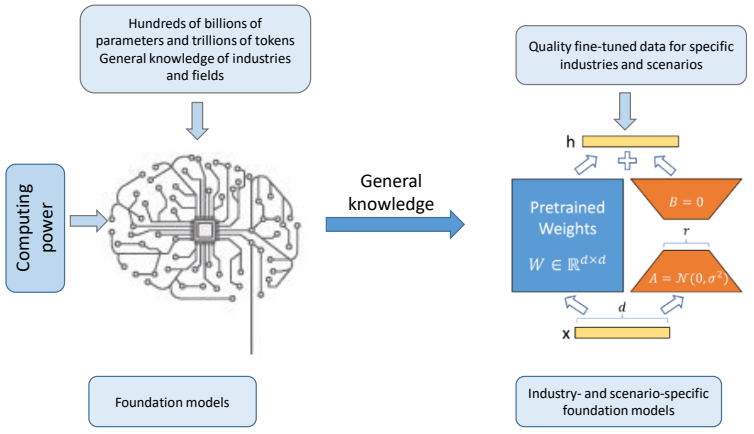
	Data Scale	Data Diversity
Small models	A dataset usually contains millions or tens of millions of data records, suitable for resource-constrained scenarios (such as mobile devices or edge devices).	The data is usually used for a specific task or domain (such as medical image classification and speech keyword recognition), and a wide coverage is not needed. The data is more centralized to optimize the model performance in specific scenarios.
Large models	Large models need to process and learn more complex relationships and patterns, and therefore require large-scale datasets. For example, the Common Crawl dataset used to train GPT-3 contains hundreds of billions of tokens, a significant increase compared with the dataset used to train a small model.	The datasets involve multiple domains and topics, including but not limited to science, technology, culture, art, history, and business. Datasets are no longer limited to a single modality or a specific domain, but present multi-modal and cross-domain characteristics.

Large Model and Small Model

Small models are light and highly efficient.
They are suitable when resources are limited and only one task needs to be performed, for example, a watch that is only used to check the time.

Large models feature higher processing capabilities and accuracy. They are suitable when high complexity and accuracy are required.
For example, a smartwatch can be used to check the time, heartbeat, sleep, and other parameters.

How Does a Foundation Model Learn?



HUAWEI

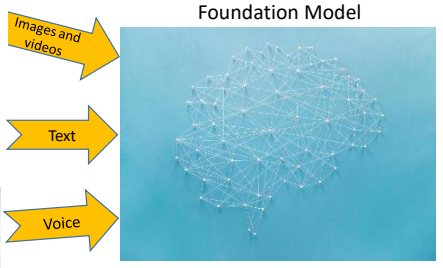
- Foundation models require massive data for learning. It is costly to train models for each industry and scenario from scratch. Therefore, the learning of foundation models is based on a basic model pre-trained on massive data and then fine-tuning is performed based on industry data.

Foundation Models That Can Speak and Paint

- Unlike conventional single-modal models (which process only one type of information such as text, images, or audio), a foundation model can integrate multi-modal data (text, image, video, and audio) for comprehensive understanding and inference.
- Cross-modal understanding: The relationship between different types of data can be understood. For example, information is extracted from an image and described in text, or an image or video is generated based on a text description.

Comprehensive model pipeline

Provides an end-to-end tool chain for model development, training, and inference, together with seamless coordination among DataOps + MLOps + DevOp, making development 50% more efficient.



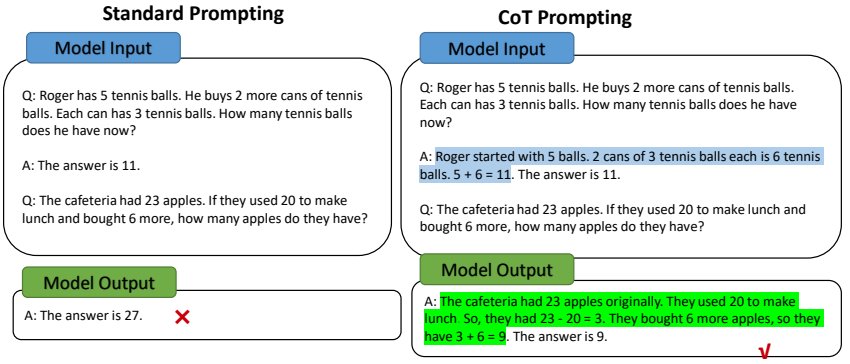
AI Breakthroughs Driven by Foundation Models - Emergence

- When a complex system consists of many micro individuals, these micro individuals are combined and interact with each other. Integrating many micro individuals can result in a special phenomenon at the macro level that cannot be explained using a single unit. This phenomenon is called "emergence".
- Currently, two task types are considered to experience emergence:
  - In-context learning (ICL): ICL helps algorithms understand the meaning and relationships among words in a sentence and recognize the relationships among different objects in an image.
  - Chain-of-thought (CoT): Users write down step-by-step inference processes and provide them for a large language model (LLM) so it can perform some complex inference tasks.

- Emergence also exists in daily life, for example, snowflake formation, traffic jams, animal migration, and vortex formation. Snowflakes are used as an example for explanation. A snowflake is composed of small water molecules. However, if a large number of water molecules interact with each other under the premise of external temperature change, a regular, symmetric, and beautiful snowflake will be formed at the macro level.
- When the model size is not large enough, tasks cannot be processed properly. However, when the model size exceeds a threshold, those tasks can be properly performed in a sudden.

Stepwise Thinking - CoT

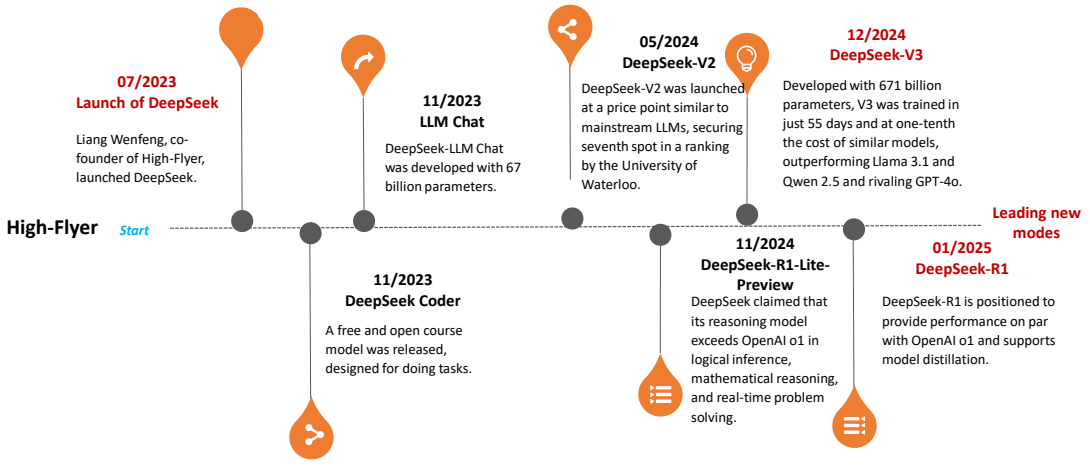
- COT is an improved prompting strategy that helps LLMs perform better in complex inference tasks, such as arithmetic, common sense, and symbolic inference. Unlike ICL, where prompts are constructed using simple input-output pairs, CoT **combines intermediate inference steps**, which can import the final output to prompts.



- Compared with the traditional in-context learning (which uses x1, y1, x2, y2, ... xtest as the input to enable the foundation model to provide the output ytest), the chain of thought introduces intermediate inference prompts.
- https://bbs.huaweicloud.com/blogs/406077.

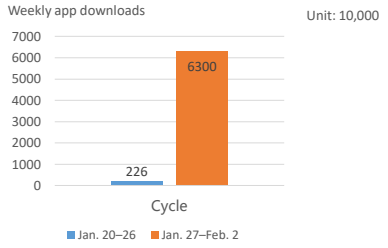
Contents

A History of DeepSeek

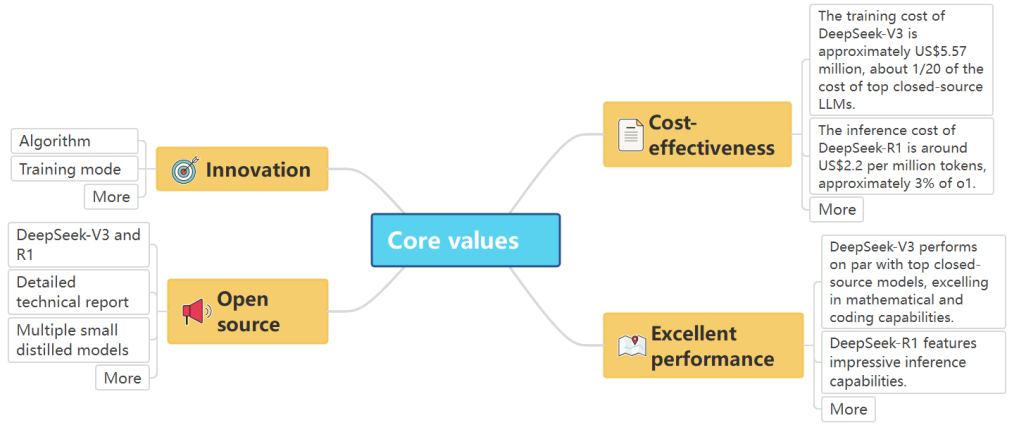


How Popular Is DeepSeek?

- On January 31, 2025, it was reported DeepSeek AI topped the list for most downloaded mobile app in 140 global markets. This led to big US companies like Microsoft, NVIDIA, and Amazon to embrace and interconnect with DeepSeek.
- As of February 9, 2025, DeepSeek had seen 110 million downloads—up from the 2.26 million downloads between January 20 and 26 and 63 million downloads the following week—with nearly 97 million weekly active users. This resulted in a week-on-week growth of 2,700%, a remarkable feat achieved without significant marketing investment.



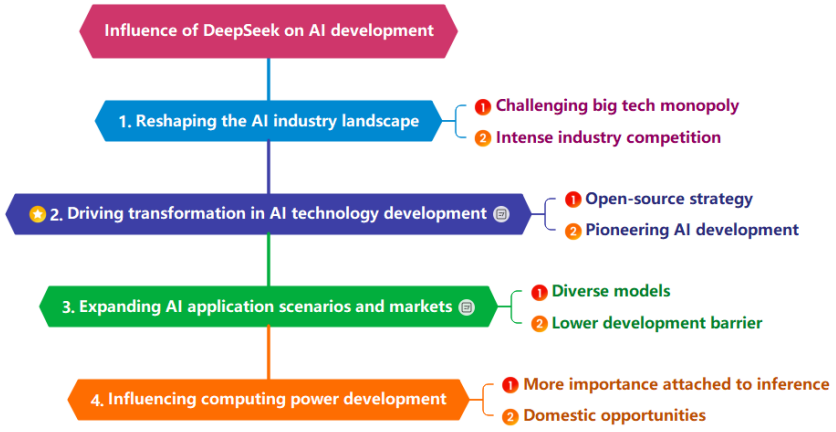
What Makes DeepSeek So Popular?



HUAWEI

- DeepSeek provides performance that rivals GPT-4o, but at a much lower training cost. While ChatGPT-4 ran up costs reaching US$78–100 million, DeepSeek-V3 was launched with just US$5.576 million (2.788 million H800 GPU hours)—just 5–10% of similar models.
- DeepSeek uses the new Mixture of Experts architecture, Multi-Head Latent Attention, auxiliary loss-free load balancing, and Multi-Token Prediction.
- V3 uses an innovative approach for training.

Influence of DeepSeek on AI Development



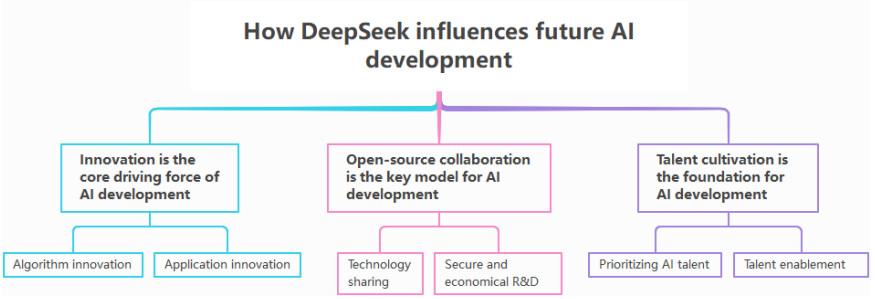
HUAWEI

- Challenging big tech monopoly: For a long time, the AI field has been dominated by a few tech giants that boast countless resources and other advantages. With DeepSeek, there is legitimate competition to the existing monopoly. DeepSeek boasts strong capabilities and a unique open-source strategy that have combined into various breakthroughs and widespread uptake in major markets. Many users and developers have moved to DeepSeek for both personal and work issues, disrupting the previous balance and prompting rival companies to reassess their market strategies.
- Intense industry competition: DeepSeek's success has inspired more companies and research institutions to invest in AI. To gain a competitive edge, major companies are investing more in R&D by introducing more cutting-edge technologies, which have not only improved technical performance but also reduced costs and expanded application scenarios. In the LLM market, this has sparked a price war, forcing other companies to lower prices to remain competitive. This greater competition has catalyzed faster technological iterations and ensured a baseline for AI development costs.
- Open-source strategy: DeepSeek is open-source, meaning it is available for anyone to download, copy, and build upon, marking a break in the typical method of closed-source AI. As part of their drive for sharing and collaboration, DeepSeek is available for developers for secondary development of new, sector-specific technologies, advancing

the application of AI in various fields.

- Pioneering AI development: Prior to DeepSeek's emergence, AI development required huge investment in both money and computing resources. This notion has changed with DeepSeek, which demonstrated that achievements can be made even under limited resource conditions through using innovative algorithms and efficient training methods. Now, AI development is not simply reliant on computing power and model scale, but on innovations in algorithms, resources, and applications, paving the way for new avenues of AI development.
- Diverse models: DeepSeek delivers powerful language understanding, text generation, mathematical reasoning, and code processing, meaning it can be applied to meet diverse needs in many different industries and domains. In finance, it can be used for risk assessment and investment decisions; in healthcare, it can assist in diagnosis and drug development; and in education, it can help develop personalized learning and intelligent tutoring. The countless benefits have seen wider support and popularity for AI in everyday life.
- Lower development barrier: Open-source models have reduced the barrier for enterprises and developers to develop AI. Instead of starting from scratch, developers instead have free tools to quickly build complex AI models or applications. Now, SMEs or other small-scale teams can develop and train applications quicker and more economically.
- More importance attached to inference: Since development costs are set to decrease with each iteration, the focus will shift to balancing cost and performance in the inference phase. This will likely lead to changes in computing power structures. The importance of inference computing power will increase significantly, driven by user needs.
- Domestic opportunities: Numerous chip manufacturers have adapted their products to DeepSeek, influencing global chip design considerations. This is advantageous for major Chinese processor companies, as it can stimulate a wider ecosystem and provide more opportunities for domestic computing power.

How DeepSeek Influences Future AI Development



- Algorithm innovation: DeepSeek achieves efficient training and powerful performance with limited resources thanks to its excellent algorithms, an enhanced Transformer architecture, and a unique attention mechanism and training framework. Since algorithms are crucial for advancing AI, companies and research institutions should increase investment accordingly, aiming to explore new paths for enhanced system performance and efficiency.
- Application innovation: Another notable aspect of DeepSeek in various fields demonstrates the potential for deep integration of AI across industries. Companies are looking to explore AI in their practices and business models to improve production and gain tangible value.
- Technology sharing: Open-source models can be shared and used by global developers, encouraging wider technology sharing and collaboration. By open-sourcing code and models, ecosystem users can work together to improve AI technology, while companies can actively participate in open-source projects to share their achievements and hear from others' experience.
- Secure and economical R&D: Open-source projects can lower R&D costs and risks. SMEs and startups can leverage open-source AI to create their products or services without

substantial investment or uncertainty from starting from scratch. Further feedback from the open-source community can help refine and enhance products for added market competitiveness.

- Prioritizing AI talent: The success of DeepSeek can be attributed to the exceptional AI team working behind the scenes. With the demand for AI growing every year, there is an equal demand for talented personnel who have a solid foundation in mathematical and computer science, expertise in core AI technologies, and a drive for innovation. Governments, universities, and enterprises should collaborate to provide initiatives that encourage the next generation of AI talent. This includes boosting investments in AI education and nurturing more professionals based on industry demands.
- Talent enablement: Enterprises and research institutions need to create a favorable environment to attract and retain excellent AI personnel. This means providing competitive compensation, development opportunities, a supportive atmosphere, and encouraging innovation and exploration. It is also important to ensure objective and fair talent evaluation that prioritizes actual achievements.

DeepSeek in the Industry

- Major companies inside and outside China are moving to use DeepSeek or integrate it with their existing applications.
- Owing to its powerful features and open-source nature, the AI has been widely adopted in many sectors, catalyzing significant technological advancements in all fields.

Henan Yunfei Technology Development Co., Ltd. partnered with DeepSeek to explore its potential in agricultural plant protection thanks to robust data analysis. This led to better capabilities in identifying patterns of and predicting the risk of disease and pest outbreaks. The AI can recommend agricultural plant protection tailored to farms and crops based on real-time field data.

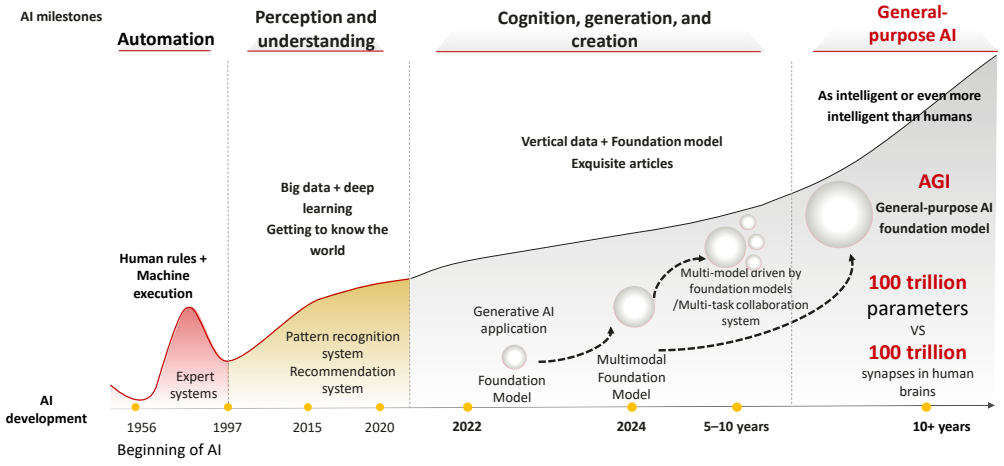
On February 5, Huawei officially linked its Celia app with DeepSeek-R1 (beta). This strategic partnership has enhanced Celia for programmers and coders, students struggling with complex math calculations, or professionals needing logical reasoning analysis. Now, users can access the DeepSeek-R1 model through the Celia app to quickly receive accurate answers.

On February 7, Geely Auto announced its self-developed Xingrui AI large model will integrate with DeepSeek.

On February 7, Dongfeng Motor announced its entire passenger vehicle lineup is integrated with DeepSeek LLMs. Wholly-owned brands it has including MHERO, eπ, Aeolus, and Nammi will soon adopt the AI-powered system.

Contents

AI Application History - From Perception and Understanding to Creation and Generation

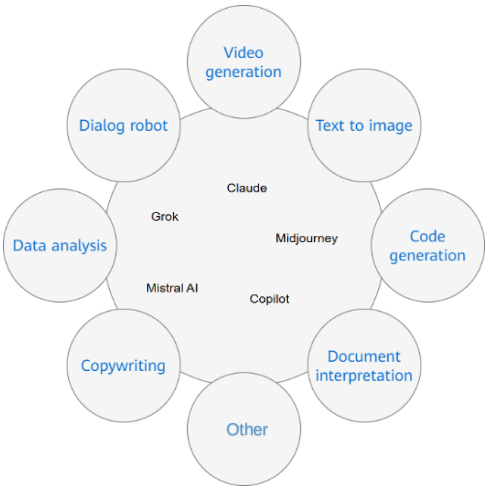


Weighty Application - ChatGPT

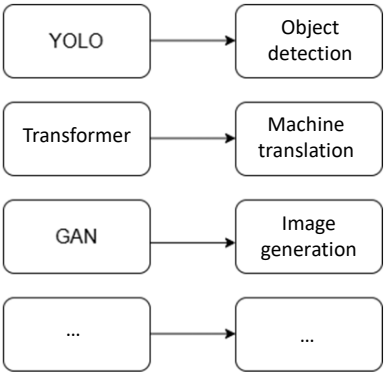
- ChatGPT is a next-generation dialog-based AI system developed by OpenAI based on its GPT language model.
  - ChatGPT had over 100 million active users only two months after its launch, making it the fastest growing consumer-level application in history.
- ChatGPT is acclaimed for human-like understanding and inference capabilities in language, context, intent, and logic. It has outstanding performance in high-quality, multi-round, and long-text human-machine interaction, and has achieved preliminary "intelligent emergence."
- ChatGPT is currently in its GPT4-o version with multi-modal capabilities and multi-domain expertise.

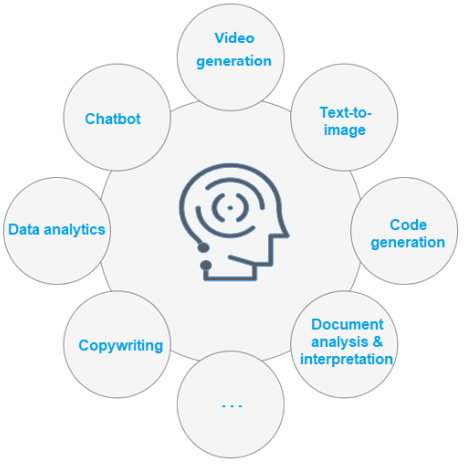
Applications Similar to ChatGPT

- After the release of ChatGPT, many similar applications have been released in China and globally.
- Today, these applications have integrated multiple functions, such as text, code, video, and image generation.

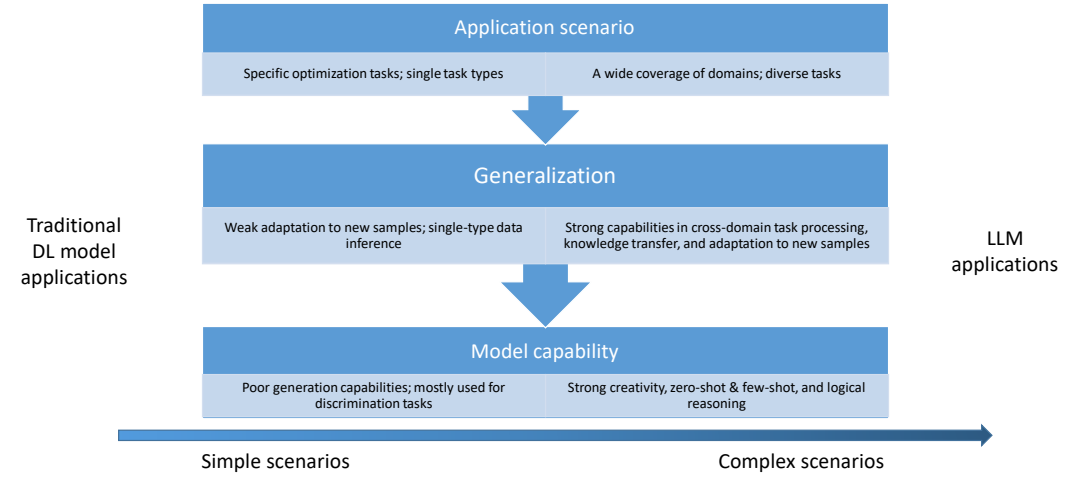


From Single Tasks to Multiple Tasks





From Simple Tasks to Complex Tasks



Contents

Intelligent Mining

Scenario 1: Electromechanical chamber - storing power distribution and communications equipment

Objective: **To implement unattended automatic inspection**

Challenge 1: 24-hour monitoring of the chamber environment

Challenge 2: Scheduled manual inspection

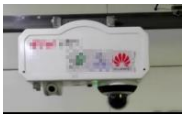
Scenario 2: Underground intelligent monitoring

	Challenges
Coal conveyor belt	It's unsafe and time-consuming to inspect a coal conveyor belt longer than **20 km**. Multiple belt segments need to collaborate to prevent coal accumulation.
Behavior protocol violations	Not wearing safety helmets, smoking, passing under equipment, walking in laneways, and sitting on belts
Operational violations	Improper operations during gas drainage and water inspection and drainage

Solution for scenario 1: Automatic inspection robot

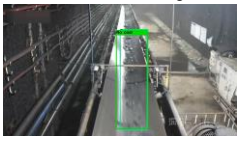
Video surveillance: **AI** machine vision for 24-hour monitoring, alarms, and sensitive area inspection

Sound detection: **AI** intelligently identifies whether machines are running properly



Scenario 2

Belt tear monitoring



Belt deviation monitoring

HUAWEI

- https://e.huawei.com/cn/material/wireless/4d008289c5424b31a055c71eaaed5790

AI Helps Protect Nature - Preserving Chile's Biodiversity

- The Nature Guardian project uses Huawei Cloud and AI to research and protect Darwin's foxes in the Nahuelbuta Mountains. The Nature Guardian is an acoustic monitoring system developed by Rainforest Connection (RFCx) and has been effective in several projects.
- It consists of solar devices equipped with microphones and antennas. These devices collect sound data from the surrounding environment and transmit it to the cloud through wireless networks for AI data analysis. Each device can cover three square kilometers around the clock.
- The Nature Guardian can capture animal sounds as well as illegal noise made by poachers' gunshots or trucks and electric saws of illegal loggers.
- The trained AI model can identify the sounds of different animals, enabling experts to study the distribution and behavior of species and helping with environmental protection through adaptive management.
- If a threat is identified, the system sends a real-time alarm to the ranger's mobile application for a fast response.

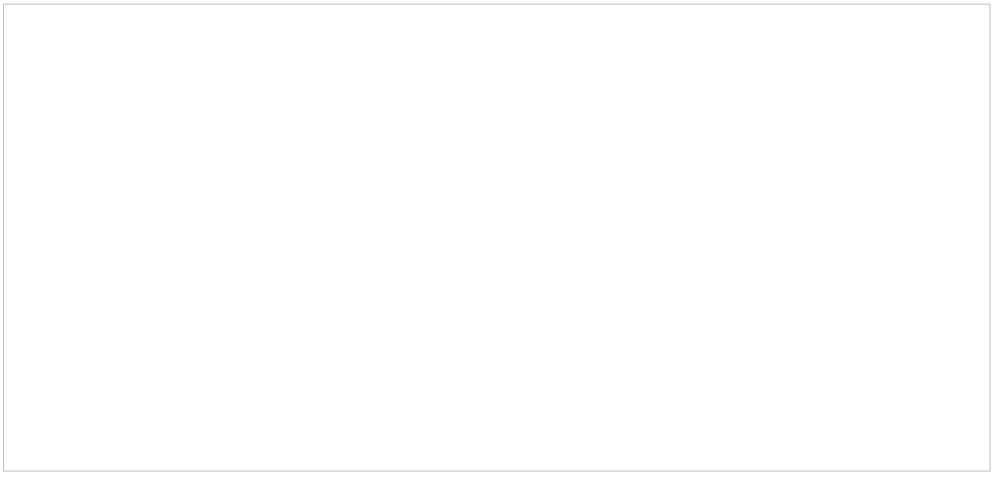
Darwin's Fox

Source:https://www.huawei.com/cn/tech4all/stories/nature-guardians-for-biodiversity-in-chile

HUAWEI

- https://www.huawei.com/cn/tech4all/stories/nature-guardians-for-biodiversity-in-chile

AI Safeguards Nature - Protecting Wildlife in Greece with a Shield of Sound



- https://www.huawei.com/cn/tech4all/stories/wildlife-greece

Scaling Law in AI for Science

Multi-disciplinary and multi-scale data to train four fundamental science models: life, materials, fluids, and electromagnetism


Atomic orbitals

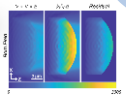

Crystal structure


Small molecule

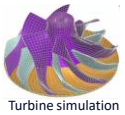

Proteins

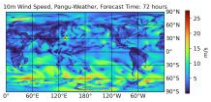

Cell


Brain


Electromagnetic simulations

Geophysical Electromagnetic Inversion

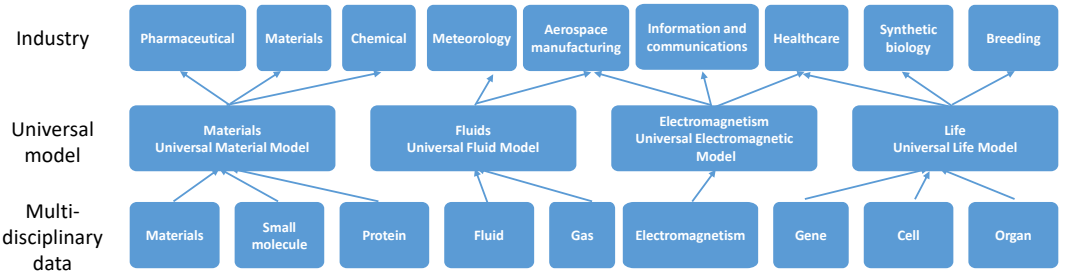

Turbine simulation


Weather forecasting

HUAWEI

AI4Science

Emergent intelligence based on massive data in various industries

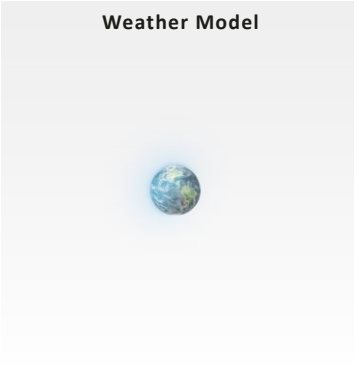


AI4Science - AI + Meteorology

Weather models for fast, accurate weather forecasting

80

typhoons worldwide each year



3D weather model | Hierarchical temporal aggregation

More accurate

Forecasting accuracy **20%** ↑

Faster and greener

Compute used for generating a 10-day weather forecast

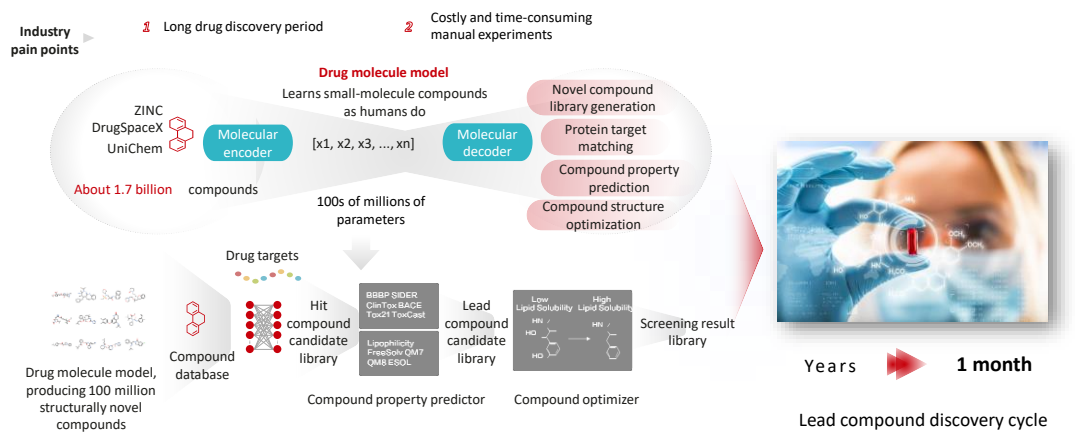
Traditional Numerical Weather Prediction (NMP)	Weather model inference
3,000 servers	1 device
5 hours	10 seconds

Forecasting speed **10,000x** ↑

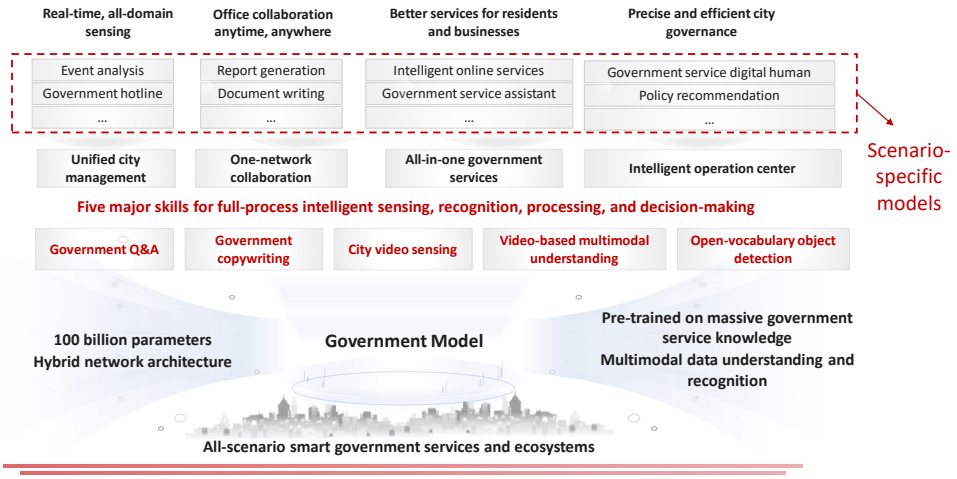
HUAWEI

AI4Science - AI + Drugs

Drug molecule models for much higher drug design efficiency



Smart Government



HUAWEI

AI For Coding

How to format a large number to two decimal places without scientific notation in ArkTS?

Solution for Formatting Large Numbers in ArkTS

Implementation Approach

1. **String-Based Processing**
   Directly manipulate numeric values as strings to prevent precision loss and avoid automatic scientific notation conversion.
2. **Scientific Notation Handling**
   Implement custom logic or integrate libraries like `bignumber.js` (recommended) for accurate scientific notation conversion.
3. **Decimal Precision Control**
   - Split the number into integer and decimal parts
   - Truncate decimals to 2 digits using `.slice(0, 2)`
   - Ensure two decimal places with `.padEnd(2, '0')`

Core Implementation

typescript

```typescript
// Formatting function
function formatLargeNumber(value: string): string {
  // Handle scientific notation (custom logic required)
  let numStr = value.includes('e') ? convertScientificNotation(value) : value;

  // Split integer and decimal parts
  const [integerPart, decimalPart = '00'] = numStr.split('.');

  // Format decimals
  const formattedDecimal = decimalPart.slice(0, 2).padEnd(2, '0');
```

3. **Library Integration**

typescript

```typescript
import BigNumber from "bignumber.js";
// Recommended for precise large number handling
const x = new BigNumber('1234567890123456789O.1234567890');
```

HarmonyOS Component Integration

typescript

```typescript
@Entry
@Component
struct DemoPage {
  @State inputValue: string = '';
  @State displayValue: string = '';

  build() {
    Column() {
      TextInput({ text: this.inputValue })
        .onChange((value: string) => {
          this.displayValue = formatLargeNumber(value);
        })
      Text(`Formatted: ${this.displayValue}`)
    }
  }
```

Best Practices

- Follow HarmonyOS performance specifications
- Use `@State` decorators for UI updates
- Validate input formats in `onChange` events
- Combine with locale-aware formatting when displaying currencies/measurements

Contents

Debates in AI

- There are many debates around AI technologies, especially as they evolve:
  - Is what we see or hear true?
  - How do we solve the ethical issues in AI?
  - Will everyone lose their jobs in the future?
  - Will our privacy be protected? How do we protect our privacy?
  - Will AI be controllable in the future?
  - ...

HUAWEI

- What you see may not be true.
- Ethical issues will be controlled by policies and regulations.
- Partial unemployment may be caused thereby.
- Privacy may possibly be leaked. However, privacy can be protected depending on technologies such as differential privacy training, model fingerprint, and model encryption.
- Controllability.

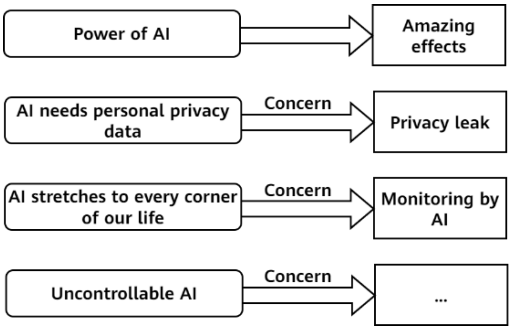
Is Seeing Still Believing?

- As AI technologies develop, we begin to question the credibility of images, audio, and video. Now, technologies such as multi-modal large models and generative adversarial network (GAN) can be used to produce false images and videos, making it difficult to distinguish between what is true and what is false.
  - For example, Lyrebird is a tool that can automatically imitate human voices from several minutes of recording samples.
  - Deepfake can generate videos with fake faces.

True?  False?

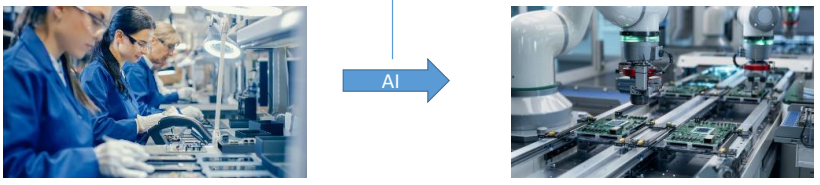
Opinions on Ethical Issues in AI

- Discrimination and stigmatization: Threatens social fairness and justice and restricts people's freedom of choice.
- Autonomous decision-making: This mechanism endangers human autonomy. Results cannot be traced back and effective measures cannot be taken to correct decisions.
- Disadvantages: The objectives, methods, and decisions of AI systems cannot be explained.
- Harm to human dignity and mind, and physical damage: People may face unsafe environments (for example, coexistence with robots) and vulnerability to the malicious use of robots, privacy leaks, and susceptibility to deception and manipulation.



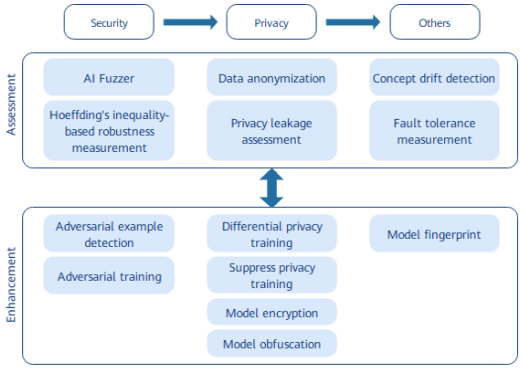
Will You Lose Your Job?

- As a society, we have always been in pursuit of higher efficiency. The steam engine reduced the need for horses. Every step in achieving automation will change our life and work. In the era of AI, what jobs will be replaced by AI?
- AI may replace:
  - Repetitive work
  - Hazardous work
  - ...



Privacy and Data Security

- Big data and AI allow us to obtain information more conveniently and efficiently. However, our behavior is also recorded and the data is used all the time. How can we protect private data?
  - 1. Users: Before obtaining the user's information, mobile applications must obtain the user's consent (**privacy policy**). Users can refuse to provide the information required by an application for proactive protection.
  - 2. Society: **Actively promote the formulation of privacy protection clauses, implement basic industry rules**, and strengthen supervision.
  - 3. Developers: **Confidential computing, model privacy protection, federated learning, or adversarial learning** can be used to prevent reverse attacks and protect user privacy. Alternatively, a well-established privacy protection framework.

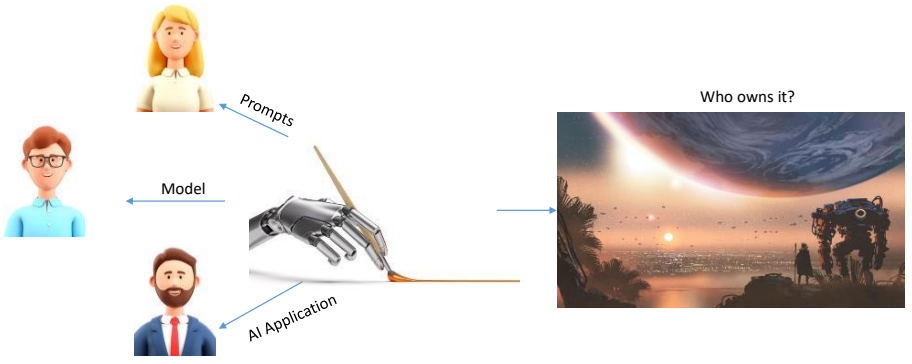


HUAWEI

- Confidential computing: Data transmission and computing are confidential, and privacy protection is costly.
- Model privacy protection: Commonly used methods include data anonymization, differential privacy training, model encryption, and model obfuscation. (Differential privacy can measure and control the leakage of trained data by models.)
- Federated learning: Generally, multi-party machine learning is performed without sharing data. It is essentially a distributed machine learning framework with limited data access. Computing nodes share only gradients but not raw data.

Problems to Be Solved

- Are AI-created works protected by copyright laws?
- Who gives authority to robots? What rights shall be given to robots?
- ...



Contents

Value - Wide Application

- AI will see wider application as vertical fields integrate the technology.
  - Smart transportation: highways, airports, railways, logistics, ports
  - Energy: electricity, oil, gas
  - Finance: banking, insurance, and securities
  - And more

- Digital government: smart city, smart government, smart emergency response, sunny kitchen, smart water conservancy, and the like. For example, sunny kitchen uses AI machine vision to identify food types on meal plates and calculate prices and calories. Smart water conservancy and management can solve problems in the following aspects: 1. black water management and sewage monitoring (AI video algorithm for sewage identification, satellite remote sensing, and water pump gate scheduling); 2. reservoir (enhanced video, intelligent power supply, and integrated pole site for water conditions) 3. rivers and lakes (spectral water quality analysis, and 5G access to drones/ships).
- Smart mining: underground 5G network, security monitoring, and violation detection.
- Smart transportation: airports (IOC queuing management – sensing an extra long passenger queue: AI + video analysis; stand scheduling: operation optimization AI algorithm + big data + IoT + GIS + video + simulation...), and highways (holographic intersection, unmanned driving...).

Value - Growing Market

Market scale

The global AI market was expected to reach USD638.2 billion in 2024 (USD538.1 billion in 2023).

Total AI IT investment

In 2022, the total global investment in AI IT was USD128.8 billion, which is estimated to rise to USD423.6 billion in 2027, representing a five-year compound annual growth rate (CAGR) of about 26.9%.

Foundation Model Solutions

In the first half of 2024, the market scale of AI foundation model solutions in China was CNY1.38 billion. The estimated CAGR is 56.2% from 2024 to 2028, reaching a market scale of CNY21.1 billion by 2028.

Value - Higher Production Efficiency

- AI innovation plays a prominent role in advancing the manufacturing industry. Smart manufacturing can significantly improve product quality and production efficiency. Statistics show that AI reconstruction helps speed up R&D in factories by 20.7% and improve production efficiency by 34.8%.
- Labor-intensive and technology-intensive enterprises should make full use of AI innovation. The former can use AI for higher labor productivity, while the latter has strong R&D capabilities and is somewhat ready for AI innovation.

A group of people work together.

A few people control a group of machines.

Value - Industry-specific AI Applications

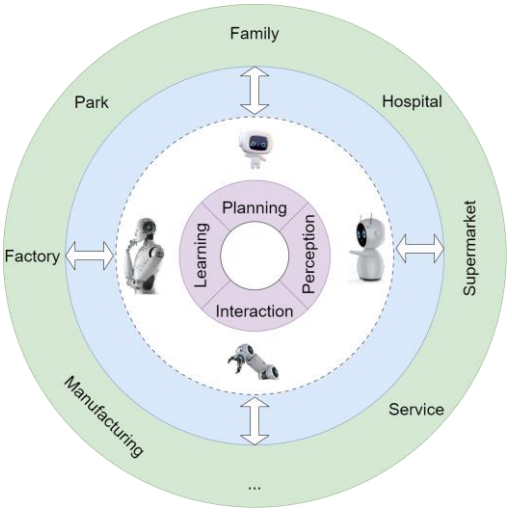
- In addition to offering general capabilities, AI technologies are also customized for vertical industries to solve pain points in specific domains and provide differentiated solutions.
- Industry-specific customization unlocks the true value of AI technologies. In the future, competition will increasingly hinge on a deep understanding of vertical scenarios.

Technology - Artificial General Intelligence

- AGI is a computer system that can implement any intelligent human activity. It features universal intelligence, including self-learning, self-improvement, self-adjustment, and cross-domain and cross-modal learning, inference, and decision-making capabilities.

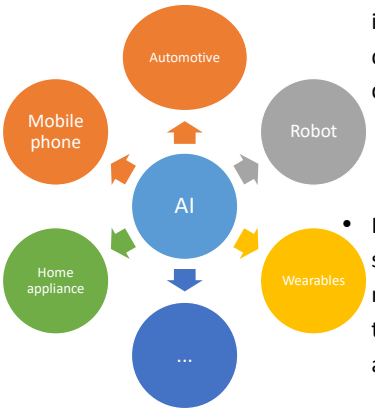
Technology - Embodied Artificial Intelligence

- Embodied artificial intelligence is an intelligent system that implants AI in physical objects. This gives objects capabilities like autonomous perception, learning, decision-making, and action in physical environments, so they can flexibly adapt to physical environments and tasks.



AI Devices

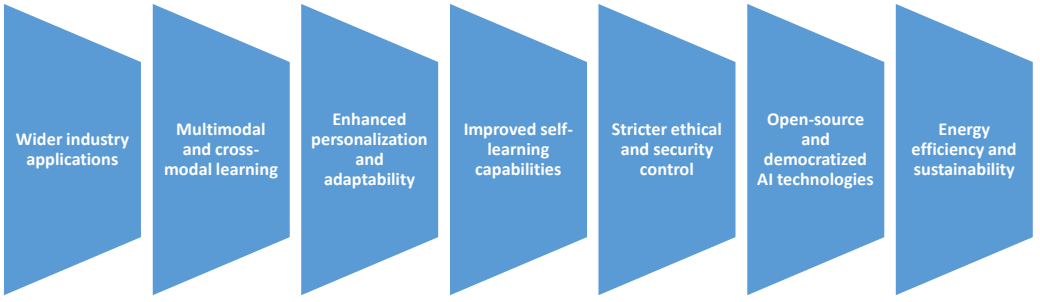
- Smartphones will integrate AI, enabling natural interactions and becoming truly smart. They will deliver a brand-new experience together with wearable devices.
- Smart home appliance will provide recipe recommendations, optimize energy usage, and offer language interaction for a more comfortable experience.



- Intelligent vehicles will achieve intelligent driving in a real sense, delivering safe and comfortable driving experiences.
- Intelligent robots will feature stronger sensing and decision-making capabilities, ready to take on jobs like housekeeping and nursing.

Future of AI Applications

- The future of AI applications is expected to see wider technology integration, deeper industry penetration, and continuous ethical and security considerations.



- Wider industry applications: As technologies become more mature and adaptable, large models are poised to find broader applications in traditional industries, such as manufacturing, energy, and logistics, to support more efficient problem-solving and decision-making.
- Multimodal and cross-modal learning: Future large models will not be limited to a single domain such as text or images. Instead, they will be able to process and integrate multiple types of data (such as text, images, videos, and sounds) to provide a more comprehensive AI application experience.
- Enhanced personalization and adaptability: Large models will better adapt to specific needs of individual users and provide customized services. For example, in the education and health fields, models can provide personalized suggestions based on individuals' learning progress or health status.
- Improved self-learning capabilities: Future large models will have stronger self-learning capabilities, allowing them to learn new patterns and knowledge through continuous interaction and feedback. This reduces the reliance on large-scale labeled data.
- Stricter ethical and security control: As large models become more widespread, ethical, privacy, and security issues will receive more attention. Future development will focus

on ensuring the transparency, explainability, and fairness of models while strengthening data protection.

- Open-source and democratized AI technologies: The development and application of large models may become more democratized. More research institutions and companies may make their models and technologies open-source to promote global technology sharing and innovation.
- Energy efficiency and sustainability: Given the enormous energy consumption required to train large models, future research will focus on improving the energy efficiency of models and reducing the carbon footprint to support sustainable development.

Summary

- This chapter covered the definition and history of AI, its applications and sectors, and controversial topics and future trends in the field.

Quiz

1. (Multi-choice question) Which of the following are applications of computer vision?
   A. Knowledge graph
   B. Semantic segmentation
   C. Intelligent driving
   D. Video analysis

- Answer:
  - BCD

Recommendations

- Huawei Talent Online Website
  - https://e.huawei.com/cn/talent/#/home
- Huawei Cloud
  - https://www.huaweicloud.com/intl/en-us/

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.