International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 1 ǁ January 2025 ǁ PP. 180-186

Background, History, Concepts, Applications, Challenges and Trends of Large Models

Jing Lei[1*], Jia-Qing Song[2], Jing Li[1], Yan Liang[1]
[1] *School of Mathematics and Statistics, Taishan University, Tai'an, China*
[2] *Experimental Teaching Management Center, Taishan University, Tai'an, China*
[*]*Corresponding Author: elizabethia@126.com*

Abstract
With the rapid development of artificial intelligence technology, large models have become the core technology in the fields of natural language processing, computer vision, and speech recognition. This article aims to comprehensively discuss the background, development history, concepts, applications, challenges, and future trends of large models. Firstly, the background of the generation of large models is introduced, and its important position in the development of artificial intelligence is expounded. Secondly, the development process of the large model is reviewed, and the key technological breakthroughs at different stages are summarized. Then, the concept and main features of the large model are explained in detail. Then, the application practice of large model in multiple fields was analyzed, and its great value in practical scenarios was demonstrated. In addition, the challenges of computing resources, data privacy, and model bias in the development of large models are discussed. Finally, the development trend of large models is prospected, including model scale, multimodal fusion, lightweight, and interpretability. The research in this paper will help deepen the understanding of large models and provide reference for researchers and practitioners in related fields.

Date of Submission: 17-01-2025 Date of acceptance: 31-01-2025

I. INTRODUCTION

In recent years, with the popularity of cloud computing, the Internet of Things, and mobile communication technologies, data has been generated at an unprecedented speed and scale. This data is not only huge, but also diverse, including text, images, sounds, and other forms. How to process and analyze these complex data, and effectively extract valuable information and knowledge from these massive data to achieve in-depth understanding and prediction of the real world, is an important topic facing the AI field at present, and has become the focus of both academia and industry. Traditional models often have problems such as small parameter scale and weak expression ability when dealing with complex tasks. In order to overcome these limitations, researchers have begun to focus on the study of large models. As an emerging AI technology, large model provides new ideas and methods to solve this problem with its powerful feature extraction and pattern recognition capabilities [1]-[4].

Researchers at home and abroad have conducted extensive and in-depth research on large models. Early research mainly focused on the construction and optimization of models, such as support vector machines (SVMs) and random forests. With the rise of deep learning technology, the research of neural network models, especially deep neural networks (DNNs), has become a hot topic. The researchers found that by increasing the depth and width of the model, the performance of the model on tasks such as image recognition, speech recognition, and natural language processing can be significantly improved.

However, as the size of models continues to expand, how to effectively train and deploy these large models has become a new challenge. Researchers have begun to pay attention to technologies such as parallel computing, distributed training, and model compression. At the same time, the application effect of large models in theory and practice has also received extensive attention. The research shows that large models have significant application potential in the fields of medical diagnosis, financial analysis, and intelligent recommendation.

The development of large models not only reflects the progress of AI technology, but also is a key factor in promoting the intelligent transformation of various fields of society. First of all, the in-depth study of large models can help improve the processing power of AI technology, so that AI systems can better understand and respond to complex and changeable data environments. Secondly, the application of large models can promote the innovation and development of various industries, improve production efficiency, and optimize resource allocation.

It can be seen that the study of large models has important theoretical value and practical significance. First of all, through the study of large models, we can promote the innovation of relevant algorithms and theories,

and provide new impetus for the development of AI technology. Secondly, exploring the application of large models in different fields will help solve practical problems and improve the technical level of society as a whole. Moreover, in-depth research on the problems encountered in the training and deployment of large models can provide technical support for the popularization and industrialization of AI technology.

Based on the background of large models, this paper reveals the necessity and significance of large model research through a review of the existing literature, discusses the relevant theories and applications of large models, discusses the unsolved problems and challenges of large models, and finally predicts the future development trend of large models.

II. HISTORY

The history of large models is a magnificent chapter in the field of artificial intelligence, which has witnessed a historic leap from simple algorithms to complex neural networks [1]-[4]. The development of large models can be divided into the following five important stages:

(1) Early exploration (1950s-1980s)

The origins of large models can be traced back to the 50s of the 20th century, when computer scientists began to explore how to simulate human intelligence through algorithms. In 1957, Frank Rosenblatt invented Perceptron, the simplest model of neural networks, marking the beginning of the study of large models. However, due to the limitations of computing power and data resources, the model of this period was relatively simple and had a limited range of applications.

(2) The Rule-Driven Era (1980s-1990s)

From the 80s to the 90s of the 20th century, artificial intelligence research entered a rule-driven era. During this period, Expert Systems and rule-based systems became mainstream. These systems rely on human-written rules to process information, and while they are automated to some extent, their limitations in the face of complex problems are becoming apparent.

(3) The Rise of Statistical Learning (1990s-2000s)

In the 90s of the 20th century, statistical learning methods began to emerge, especially the emergence of algorithms such as Support Vector Machine (SVM) and Random Forest, which provided new tools for processing large-scale data. These methods greatly improve the efficiency and generalization of machine learning by learning patterns from data rather than relying on predefined rules.

(4) Deep Learning Revolution (2006-present)

In 2006, Geoffrey Hinton and others reintroduced the concept of deep learning and promoted the development of deep neural networks (DNNs). In 2012, AlexNet achieved a breakthrough in the ImageNet competition, marking the dawn of the era of deep learning. Subsequently, a series of deep learning models such as VGG, GoogLeNet, and ResNet have been developed one after another, constantly refreshing the performance records of tasks such as image recognition.

(5) The era of large models (2018-present)

In 2018, Google launched the Transformer model, which has made remarkable achievements in the field of natural language processing (NLP), especially the emergence of its variant BERT (Bidirectional Encoder Representations from Transformers), which opened the era of large models in the field of NLP. Subsequently, OpenAI's GPT series models, Microsoft's Turing NLG and other super-large pre-trained models have been released, which have billions or even hundreds of billions of parameters and can show excellent performance on multiple tasks.

Factors that play a key role in the development of large models include:

- Improvement of computing power: With the development of special computing hardware such as GPU and TPU, the training of large models is possible.
- Growth of data volume: The popularization of the Internet and the development of big data technology provide rich data resources for training large models.
- Algorithm innovation: From perceptron to deep learning, the continuous innovation of algorithms has promoted the development of large models.
- Driven by application demand: The growing demand for intelligent solutions has promoted the application of large models in various fields.

The history of the development of large models is a history of continuous innovation and breakthroughs, which not only reflects the progress of artificial intelligence technology, but also brings revolutionary changes to all walks of life.

III. CONCEPTS

Large models are an important concept in the field of artificial intelligence in recent years, which refers to machine learning models with a large number of parameters and deep structure. These models are typically trained on large-scale datasets and are able to capture complex patterns and high-level abstract features of the data. The following is a detailed explanation of the concept of large models.

(1) Definitions

Large models typically refer to deep learning models that have hundreds to trillions of parameters. These models include, but are not limited to, deep neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), Transformer architectures, and more. The "bigness" of a large model is mainly reflected in two aspects: first, the complexity of the model structure, that is, the number of layers and neurons of the model; The second is the number of model parameters, that is, the number of variables that the model needs to learn during training.

(2) Features

1) Data-driven: Large models rely on a large amount of data for training, by learning features and rules from the data, rather than relying on artificially designed features.

2) Pre-training and fine-tuning: Large models usually adopt a two-stage training strategy of pre-training and fine-tuning. In the pre-training phase, the model learns general knowledge on a large-scale dataset; In the fine-tuning phase, the model adapts to a task-specific dataset to suit a specific application.

3) Transfer learning: Through transfer learning capabilities, large models transfer the knowledge learned on one task to other related tasks, so as to improve the generalization ability and efficiency of the model.

4) Powerful representation ability: Large models can learn the deep representations of data, which can often capture the high-level abstract features in the data, which has significant advantages for the processing of complex tasks.

(3) Type

1) Natural language processing models: such as BERT, GPT, XLNet, etc., which perform well in tasks such as language understanding, text generation, and machine translation.

2) Computer vision models: such as ResNet, EfficientNet, ViT (Vision Transformer), etc., which have high performance in vision tasks such as image recognition, object detection, and image segmentation.

3) Multimodal models: such as CLIP, Florence, etc., these models can process and understand data of different modalities, such as text, images, sounds, etc., to achieve cross-modal learning and application.

(4) Training and deployment

The training of large models usually requires a large amount of computing resources, including high-performance GPUs or TPU clusters. In addition, the training process also needs to be accelerated by efficient network communication and parallel computing technology. In terms of deployment, large models can be carried out through cloud services, edge computing, etc., to meet the needs of different application scenarios.

Large model is an advanced technology in the field of artificial intelligence, which realizes a deep understanding of complex data patterns and the capture of high-level abstract features through large-scale dataset training. With the continuous advancement of technology, large models have shown great potential and value in many fields.

IV. APPLICATIONS

As a powerful machine learning tool, large models have shown their wide application potential in many fields [5]-[11]. Here are just a few of the main application areas and their specific application examples.

4.1 Applications

(1) Natural language processing (NLP): Natural language processing is one of the most widely used fields for large models. The specific application scenarios are:

- Language models: such as OpenAI's GPT-3, which can generate coherent and logical text for writing, programming, chatbots, etc.
- Machine translation: For example, Google's Transformer model significantly improves the accuracy and fluency of translation, and supports real-time translation between multiple languages.
- Sentiment analysis: Large models can accurately identify and analyze sentiment tendencies in texts, and are widely used in market research, social media analysis and other fields.
- Text summarization: Large models are able to extract key information from long articles and generate concise summaries for news aggregation, research summaries, etc.

(2) Computer vision: In the field of computer vision, large models have also made remarkable achievements, especially in the following aspects:

- Image recognition: For example, the ResNet model has achieved a breakthrough accuracy rate in the ImageNet competition, and is widely used in image search, security monitoring, etc.
- Object detection: The large model can accurately identify multiple objects in the image and locate their positions for autonomous driving, drone monitoring, etc.
- Image segmentation: The large model can classify each pixel in the image, which is used for medical image analysis, autonomous driving map generation, etc.

(3) Speech recognition and synthesis: The applications of large models in the field of speech include:

- Speech recognition: such as the DeepSpeech model, which can convert speech to text and apply it to voice assistants, meeting records, etc.
- Speech synthesis: Large models are able to generate natural and smooth speech for voice assistants, audiobooks, movie dubbing, etc.

4.2 Application Examples

(1) Healthcare: In the field of healthcare, the application of large models is changing the face of medical services [5]. For example, in terms of disease diagnosis, large models can analyze medical images to assist in the diagnosis of diseases, such as cancer, fractures, etc. In drug discovery, by analyzing large amounts of bioinformatics data, large models can help accelerate the discovery and development of new drugs.

(2) Education: In the field of education, the application of large models is promoting the reform of teaching models [9]. For example, according to students' learning habits and abilities, large models can provide customized learning content and paths. The large model can simulate the teacher to answer questions and provide personalized learning guidance.

(3) Finance: In the financial field, large models are more widely used [11]. For example, in terms of credit assessment, by analyzing users' trading behavior and related information, large models can more accurately assess credit risk. In terms of market forecasting, large models are able to analyze market trends and news events to support investment decisions.

Although large models have shown great potential in various fields, they also face a series of challenges in their application.

V. PROBLEMS AND CHALLENGES

5.1 Unresolved Issues

Although large models have made significant progress and wide application in the field of artificial intelligence, there are still a series of unsolved problems that hinder the further development and popularization of models. These include:

(1) Calculate resource requirements

Large models typically require a lot of computing resources for training and inference. This not only leads to high economic costs, but also energy consumption and environmental problems. At present, how to effectively reduce the computing resource requirements of large models and achieve efficient training and deployment is a problem that has not been completely solved.

(2) Data privacy and security

The training of large models often relies on massive amounts of data, which raises concerns about data privacy and security. How to effectively use data to train the model while ensuring user privacy, and how to protect the model from malicious attacks and abuse are problems that need to be solved urgently.

(3) Model bias and fairness

Large models may absorb and amplify bias in the data during training, leading to discriminatory decisions when applying the model. How to ensure the fairness of the model and eliminate bias is an important ethical issue in the field of artificial intelligence.

(4) Explainability and transparency

Large models are often considered "black box" models with a lack of transparency and explainability in their decision-making process. This has become a serious problem in certain areas that require a high degree of transparency (e.g., healthcare, finance). How to improve the interpretability of large models and make their decision-making process more transparent is a hot topic in current research.

(5) The generalization ability of the model

Although large models excel at specific tasks, their generalization capabilities are often limited. It is a challenging problem to improve the generalization ability of the model in different fields and tasks, so that it can better adapt to diverse application scenarios.

(6) Model lightweight and miniaturization

In order to deploy large models in mobile devices and edge computing environments, it is necessary to reduce the weight and miniaturization of models. How to reduce the number of parameters and computational complexity of the model without sacrificing too much performance is a technical problem.

(7) The diversity and quality of training data

The performance of large models is highly dependent on the quality and diversity of the training data. How to obtain high-quality and diverse training data, and how to deal with noise and outliers in the data, is the key to improving model performance.

Unsolved problems are an important direction of large-scale model research and development, and solving these problems requires not only technological innovation, but also interdisciplinary cooperation and policy support.

5.2 Challenges

Although the development and application of large models have brought revolutionary progress to the field of artificial intelligence, it also comes with a series of unique challenges [12].

(1) Training stability

The training process of large models is often very complex, and the number of model parameters is huge, which leads to problems such as gradient vanishing or explosion during the training process. How to design a stable training algorithm to ensure that the model can learn effectively is a technical challenge.

(2) Optimization algorithms

Due to the complexity of large models, traditional optimization algorithms may not be able to train them effectively. Researchers need to develop new optimization techniques to improve training efficiency and reduce training time while maintaining model performance.

(3) Knowledge transfer

Large models excel at specific tasks, but it is a challenge to transfer what they learn to other related tasks. This involves research in the fields of transfer learning and multi-task learning, and the problem of knowledge extraction and reuse needs to be solved.

(4) Model evaluation

Evaluating the performance of large models is a complex issue. Traditional evaluation metrics may not fully reflect the capabilities of large models, especially for some highly subjective tasks. The development of more effective evaluation methods and standards is an important direction of large model research.

(5) Resource coordination

The training and deployment of large models requires a large amount of computing resources, and how to efficiently coordinate and allocate these resources to achieve optimal utilization of resources is a management challenge.

(6) Legal and ethical issues

The application of large models involves legal and ethical issues such as data privacy, intellectual property rights, and algorithmic bias. How to use large models under the premise of complying with laws, regulations and ethical standards is a common concern of all sectors of society.

(7) Model maintenance and updating

Large models need to be regularly maintained and updated to maintain their performance. However, the complexity of the model makes this process difficult. How to efficiently maintain and update the model is a practical problem after the model is deployed.

(8) User acceptance

The application of large models may change traditional workflows and user habits. How to improve users' acceptance and trust in large models and promote their popularization in practical applications is a social and technical challenge.

The challenges of large models are multifaceted and involve multiple levels such as technology, management, law, and ethics. These challenges need to be addressed through a variety of means, including interdisciplinary collaboration, technological innovation, policy development and public education.

VI. TRENDS

As the core technology in the field of artificial intelligence, the development trend of large models reflects the future trend of the entire AI field. Here are a few of the main trends in the development of large models:

(1) The scale of the model continues to grow

As computing power increases and training data increases, the size of large models will continue to grow. Larger models are able to capture more complex patterns and richer knowledge to achieve breakthrough performance. For example, the number of parameters of GPT-3 has reached the level of 100 billion, and there may be larger-scale models in the future.

(2) Multimodal fusion

Currently, large models focus on data processing in a single modality, such as text, image, or speech. The future trend will be multimodal fusion, that is, large models that can process and understand multiple types of data. This model is able to better simulate the multi-sensory experience of humans and improve the processing power of cross-domain tasks.

(3) Model lightweight

While large models have performance advantages, their large size and computing requirements limit their application in mobile devices and edge computing scenarios. Therefore, the lightweight of models will become an important trend, including model compression, knowledge distillation and other technologies will be more widely studied and applied in the future.

(4) Explainability and transparency

With the application of large models in key areas, the interpretability and transparency of models are becoming increasingly important. In the future, researchers will devote themselves to developing new methods and technologies to improve the interpretability of large models and enhance users' trust in models.

(5) Ethics and legal compliance

With the deepening of the application of large models, ethical and legal issues will become more prominent. Future trends will include embedding ethical and legal compliance considerations in model design and applications to ensure the healthy development of AI technologies.

(6) Interdisciplinary collaboration

The development of large models will increasingly rely on interdisciplinary collaboration. Computer scientists, data scientists, psychologists, sociologists, and legal experts will work together to promote the technological innovation and application of large models.

(7) Self-supervised learning

Self-supervised learning is a method for training models without labeling data, making it ideal for large-scale datasets. In the future, the self-supervised learning technology of large models will be further developed to reduce the dependence on a large amount of labeled data.

(8) Application of reinforcement learning

The application of reinforcement learning in large models will become more and more extensive, especially in games, robot control, and other fields. The powerful representation ability of large models combined with the decision-making ability of reinforcement learning will provide more effective solutions for complex tasks.

The development trend of large models shows that AI technology is developing in the direction of being more efficient, smarter, more explainable, and more compliant. As these trends continue to advance, large models will have a more far-reaching impact on human society in the future.

VII. CONCLUSIONS

This paper comprehensively discusses the background, development history, concepts, applications, challenges, and trends of large models. The emergence of large models is an important breakthrough in the field of artificial intelligence, which realizes a deep understanding of complex data patterns and the capture of high-level abstract features through large-scale parameters and deep network structures, and its emergence marks a new stage of development of AI technology. Large models have undergone an evolution process from simple to complex, from single to diversified, and with the improvement of computing power and algorithm innovation, the scale and performance of large models have been significantly improved. At present, large models have achieved remarkable results in many fields such as natural language processing, computer vision, and speech recognition, and have shown great potential in practical scenarios such as recommender systems, financial analysis, and medical diagnosis. However, despite the remarkable achievements of large models, they still face many challenges in their development, including computing resource requirements, data privacy and security, model bias, explainability, etc., which need to be addressed through technological innovation, policy development, and interdisciplinary collaboration. In the future, the development trend of large models will continue to move towards the growth of model scale, multimodal fusion, model lightweight, interpretability improvement, ethics and legal compliance, etc. These trends bode well for large models to play a greater role in more areas.

To sum up, as an important technology in the field of artificial intelligence, large models have broad development prospects, but we also need to face up to and solve the challenges they bring. With the deepening of research and technological progress, large models are expected to bring more efficient, intelligent and reliable AI services to human society, and promote innovation and development in all fields of society.

REFERENCES

[1]. Ding P., Liu Z. and Wang J. (2020) “Overview and Prospects of Multimodal Artificial Intelligence Large Models” Radio Engineering, pp.1-20.

[2]. Cai R., Ge J., Sun Z. and etc. (2024) “Overview of the Development of AI Pre trained Large Models” Small Microcomputer Systems, Vol. 45, No. 10, Pp. 2327-2337.

[3]. Luo J., Sun Y., Qian Z. and etc. (2023) “Overview and Prospects of Artificial Intelligence Large Models” Radio Engineering, Vol. 53, No. 11, Pp. 2461-2472.

[4]. Zhang Q. (2023) “Overview of the Development of AI Large Models” Communication Technology, Vol. 56, No. 3, Pp. 255-262.

[5]. Jiang H. (2025) “JD Health launches industry's first comprehensive medical and health application model” China Business Times, January 20, Pp. 4.

[6]. Mao Z., Liao X. and Ma H. (2020) "Database, Knowledge Base, and Large Model for Biomanufacturing" Journal of Bioengineering, Pp.1-21.
[7]. Tang Y. and Wang Y. (2025) "Research and Practice on Comprehensive Capability Evaluation of AI Large Models in Consulting Business" Communication and Information Technology, Vol. 1, Pp.133-136.
[8]. Zhang C. (2025) "Vigorously developing AI industry models" United Daily News, January 22, Pp. 7.
[9]. Xie Y., Zhang Y., Xu K. and etc. (2025) "Multi modal Large Model Empowering Teacher Digital Portrait Construction and Application" Open Education Research, Vol. 31, No. 1, Pp. 100-109.
[10]. Ye Q. (2025) "Tencent Angel Platform: Building a Big Model Super Pipeline" Science and Technology Daily, January 20, Pp. 6.
[11]. Liu Anping, Jin Xin, Hu Guoqiang Overview of Artificial Intelligence Large Models and Prospects for Financial Applications [J]. Artificial Intelligence, 2023, (02):29-40.
[12]. Li X., Yang Z., Huang C. and et al. (2025) "6G Endogenous Intelligent Wireless Large Model: Security, Privacy, Ethics, and High Energy Efficiency" Mobile Communications, Vol. 49, No. 1, Pp. 59-66.