Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3035

DEEPSEEK LARGE - SCALE MODEL: TECHNICAL ANALYSIS AND DEVELOPMENT PROSPECT

HaiLong Liao
School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.
Corresponding Email: Jnhailong@126.com

Abstract: This paper deeply analyzes the DeepSeek large - scale model, comprehensively elaborating on its technical architecture, training mechanism, performance, application fields, as well as the challenges it faces and future development directions. Through the research on the DeepSeek series of models, it reveals their innovations and important values in the field of artificial intelligence. The research shows that the DeepSeek large - scale model, with its unique technical advantages, demonstrates excellent performance in tasks such as natural language processing, code generation, and multimodal understanding. It provides new ideas and methods for promoting the development and application of artificial intelligence technology.
Keywords: DeepSeek; Large language model; Artificial intelligence; Multimodality

1 INTRODUCTION

In the field of artificial intelligence, the development of large - scale models has profoundly changed many research directions and application scenarios in natural language processing, computer vision, and other areas. OpenAI's ChatGPT, with its powerful natural language interaction capabilities, has set off a global upsurge in AI applications since its launch. It has become a capable assistant for people to obtain information and complete tasks in daily life and work. In the highly competitive track of large - scale models, DeepSeek has emerged as a new force, attracting global attention.

DeepSeek was founded by Liang Wenfeng, an entrepreneur born in the 1980s from Guangdong Province. He has a unique growth trajectory and innovative ideas. In 2015, Liang Wenfeng and two classmates from Zhejiang University jointly founded the quantitative hedge fund High - Flyer. By using mathematical and artificial intelligence strategies for investment, by 2019, the fund's managed assets exceeded 10 billion yuan. During the process of managing the fund, Liang Wenfeng had a deeper understanding of the potential of artificial intelligence and gradually shifted his focus to the AI research and development field.

In 2023, Liang Wenfeng founded DeepSeek with the aim of developing general artificial intelligence (AGI) comparable to human intelligence. Different from traditional technology entrepreneurs, when forming the team, Liang Wenfeng boldly recruited Ph.D. students from top domestic universities. Although these young talents lacked industry experience, they had achieved many academic research results. Under Liang Wenfeng's unique bottom - up management model, the team's creativity was fully unleashed, laying a talent foundation for DeepSeek's technological innovation.

Since its establishment, DeepSeek has developed rapidly. The DeepSeek - V3 model, released at the end of 2024, shocked the industry. It only used 2,048 NVIDIA H800 chips and had a training cost of less than 6 million US dollars, but could achieve performance comparable to that of models developed by international big companies. The DeepSeek - R1 inference model, launched in January 2025, attracted widespread attention globally. Its application quickly climbed the download rankings in the Apple App Store, once surpassing well - known applications such as ChatGPT, demonstrating strong market competitiveness.

Compared with ChatGPT, DeepSeek shows unique advantages in many aspects. In terms of knowledge timeliness, DeepSeek's training data is updated to the fourth quarter of 2023, enabling it to better capture emerging technology trends. In terms of professional field depth, DeepSeek has constructed special knowledge graphs in vertical fields such as quantitative finance, semiconductor industry chain analysis, and cutting - edge biomedicine, providing more accurate services for professionals. In complex reasoning tasks, DeepSeek's logical reasoning and mathematical proof capabilities are more outstanding. In Chinese language processing, whether it is classical Chinese translation or industry - term understanding, DeepSeek performs more proficiently. Of course, DeepSeek also has some shortcomings. Currently, its multimodal capabilities are still under development, while ChatGPT has integrated image generation and voice interaction modules and performs more evenly in general scenarios.

With the rise of DeepSeek, its influence in the field of artificial intelligence is increasing day by day. In - depth research on the technical principles, training mechanisms, and application scenarios of the DeepSeek large - scale model not only helps us understand the key factors behind its success but also provides valuable reference for the further development of artificial intelligence technology, promoting the continuous progress of this field.

2 TECHNICAL ARCHITECTURE

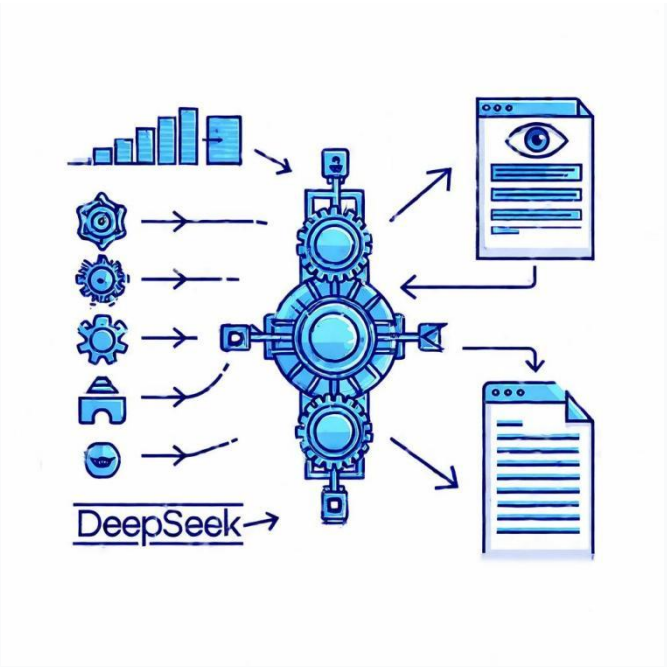

Figure 1 DeepSeek Technical Architecture Model

In the center of the figure is a large Transformer structure icon, representing the core architecture of DeepSeek (Figure 1). On the left side of the Transformer structure icon, arrows point to a series of optimization symbols (such as gears or adjustment knobs), indicating improvements to the attention mechanism to enhance the ability to process long - sequence data. In the upper - left corner of the Transformer structure icon, there is a sequence of gradually increasing number icons, symbolizing the parameter scale from small to extremely large (for example, from less than 1 billion to over 670 billion), indicating that the DeepSeek series of models have a huge number of parameters. On the right side of the Transformer structure icon, there are two interconnected small icons: one is a combination of an eye and a text box, representing visual information; the other is a simple text page icon, representing language information. These two icons are connected by a two - way arrow, indicating DeepSeek's achievements in multimodal fusion technology, especially how to effectively combine visual and language information. [1]

2.1 Innovation Based on the Transformer Architecture

The DeepSeek large - scale model is based on the Transformer architecture. The Transformer architecture, based on the attention mechanism, can effectively process sequence data and has achieved great success in natural language processing and other fields. [2]

DeepSeek has made innovative improvements to the Transformer architecture to enhance the model's ability to process long - sequence data.[7] By optimizing the attention mechanism, the model can more accurately capture the dependency relationships between various parts of the text. Thus, when processing long texts, it can have a deeper understanding of semantics and improve the accuracy and logic of the generated text.

2.2 Huge Model Parameters and Scale

The DeepSeek series of models have a huge parameter scale. For example, the DeepSeek - V3 model has as many as 671 billion parameters [3]. A large - scale parameter configuration endows the model with a stronger representation ability, enabling it to learn more abundant knowledge and language patterns. In natural language generation tasks, the model can generate more natural, fluent, and logical text. In code generation tasks, it can generate more accurate and efficient code. There is a positive correlation between the model parameter scale and performance. As the parameter scale increases, the model's performance in various tasks is also significantly improved.

2.3 Multimodal Fusion Technology

DeepSeek has made remarkable progress in multimodal fusion and has developed models such as the DeepSeek - VL2 vision - language model and the Janus - Pro multimodal model [4]. These models achieve the effective fusion of visual and language information through ingenious designs. For example, the DeepSeek - VL2 model uses a hybrid visual encoder, which can efficiently process high - resolution images (1024x1024) within a fixed token budget while maintaining relatively low computational costs. This enables the model to perform well in tasks such as visual question - answering and image description. The Janus - Pro model further enhances multimodal understanding and visual generation capabilities by decoupling visual encoding for multimodal understanding and visual generation, optimizing

the training strategy, expanding the data, and increasing the model scale[5]. In text - to - image generation tasks, it can generate high - quality images that conform to semantics according to text instructions.

3 TRAINING MECHANISM

3.1 Pretraining: Preliminary Knowledge Accumulation

In the pretraining stage, the DeepSeek model is trained using massive corpus data. These data come from a wide range of sources, including Internet texts, academic literature, code libraries, etc. Through learning on large - scale data, the model can master rich language knowledge, semantic information, and general knowledge. For example, in the training of code generation models, a dataset containing a large amount of code is used, enabling the model to learn the syntax and programming patterns of multiple programming languages. Pretraining enables the model to have basic language understanding and generation capabilities, laying a good foundation for subsequent fine - tuning.

3.2 Supervised Fine - Tuning and Reinforcement Learning: Optimizing Model Behavior

In the supervised fine - tuning stage, the model is fine - tuned on the instruction dataset. Each sample in the dataset consists of an "Instruction Q - Response A" pair. In this way, the model can better follow human instructions. In the reinforcement learning stage, DeepSeek has adopted a variety of innovative methods. Taking DeepSeek - R1 as an example, it uses a rule - based reinforcement learning method (Group Relative Policy Optimization, GRPO). In mathematical problems, a reward score is calculated based on the accuracy of the answer. In code - related problems, the compiler is used to generate feedback based on predefined test cases. At the same time, a format reward is used to ensure that the model outputs in a specific format. This method simplifies the training process, reduces costs, and enables the model to perform well in inference tasks.

3.3 Training Optimization Strategies: Improving Efficiency and Performance

DeepSeek has adopted a variety of optimization strategies during the training process to improve training efficiency and model performance. In terms of hardware resource utilization, the DeepSeek - V3 model only uses 2,048 GPUs for 2 months of pretraining, greatly reducing the training cost [3]. In terms of algorithm optimization, methods such as adjusting the learning rate and optimizing gradient calculation are used to make the model training more stable and efficient. For example, a dynamic learning rate adjustment strategy is adopted. In the initial stage of training, the model can converge quickly, and in the later stage, fine - tuning is carried out to improve the model's accuracy. [3]

4 PERFORMANCE

4.1 Natural Language Processing Tasks

The DeepSeek model performs outstandingly in natural language processing tasks. In language generation tasks such as text continuation and story creation, the generated text is coherent, logical, and has few grammatical errors. In machine translation tasks, the translation accuracy and fluency reach a high level. In the GLUE (General Language Understanding Evaluation) benchmark test, the DeepSeek model achieved excellent results, demonstrating its strong language understanding ability. This benchmark test includes a variety of natural language understanding tasks such as text entailment and sentiment analysis. The model's comprehensive performance in these tasks reflects its in - depth understanding and processing ability of language.

4.2 Code Generation and Programming

DeepSeek's code generation model has achieved advanced performance among open - source code models in multiple programming languages and various benchmark tests. It can generate high - quality code based on natural language descriptions and performs well in tasks such as code completion and code error correction. In the CodeXGLUE code generation benchmark test, the DeepSeek - Coder model is superior to many similar models in terms of the accuracy and functionality of the generated code. For a given programming problem description, the model can quickly generate correct and runnable code, effectively improving software development efficiency.

4.3 Multimodal Tasks

DeepSeek's multimodal models demonstrate excellent capabilities in multimodal tasks. In visual question - answering tasks, the DeepSeek - VL2 model can accurately understand the image content and answer related questions. [4] For an image containing multiple objects, when asked "What is the red object in the image?", the model can accurately identify and answer. In text - to - image generation tasks, the Janus - Pro model can generate high - quality images that conform to semantics according to text instructions. When inputting "Generate an image of several seagulls flying on a sunny beach", the generated image can well reflect the scene described in the text, with rich image details and coordinated colors.

4.4 Comparison with Other Models

Table 1 Comparison of DeepSeek and Other Well-known Models

Comparison Dimension	DeepSeek - V3	Other Well - known Models (e.g., Claude - 3.5 - Sonnet - 1022)
Performance	Performs close to the current excellent level in knowledge - based tasks (such as MMLU, MMLU - pro, GPQA, SimpleQA); surpasses other open - source and closed - source models in math competitions (such as AIME2024, CNMO2024).	Performs well, especially in knowledge - based tasks, being comparable to DeepSeek - V3, but may be inferior to DeepSeek - V3 in math competition tasks.
Cost - Efficiency	Low training cost, only using 2,048 GPUs for 2 months of training, costing about $5.576 million.	May require more computing resources and higher costs.
Generation Speed	The output speed has increased from 20 tokens per second to 60 tokens per second, providing a smoother user experience.	Not specifically mentioned, but it is implied that it may not be as smooth as DeepSeek - V3.

Compared with other well - known large - scale models, the DeepSeek model is competitive in terms of performance and cost - efficiency (Table 1).

5 APPLICATION FIELDS

5.1 Intelligent Customer Service and Dialogue Systems

In the field of intelligent customer service, the DeepSeek model has been widely applied. Many enterprises have integrated it into their customer service systems. The model can quickly and accurately understand user questions and provide corresponding answers. In e - commerce customer service, when users ask questions about product information, logistics status, etc., the model can respond quickly and provide accurate answers, improving customer service efficiency and user satisfaction. In dialogue systems, the model can conduct natural and smooth conversations, understand the context, and achieve multi - turn conversations, providing users with an intelligent interaction experience.

5.2 Code Development and Programming Assistance

In code development, DeepSeek's code generation model provides powerful programming assistance capabilities for programmers. It can generate code snippets based on natural language descriptions, helping programmers quickly implement functions. When developing a web application, a programmer can input "Generate the front - end code for a user login interface", and the model can generate the corresponding HTML, CSS, and JavaScript code, reducing development time and workload. The model can also perform code completion and code error correction, improving code quality and development efficiency.

5.3 Multimodal Content Creation

In the field of multimodal content creation, DeepSeek's multimodal models play an important role. In advertising design, designers can input text descriptions, such as "Design a poster to promote new energy vehicles, highlighting environmental protection and a sense of technology". The Janus - Pro model can generate corresponding images, providing creative inspiration and visual references for designers. In film and television production, the model can generate scene concept maps based on script descriptions, helping directors and art teams better conceive scene layouts. In the education field, based on image - based learning materials, when students ask questions about the image content, the model can understand the questions and combine image information to give accurate answers, assisting in teaching and students' independent learning.

6 CHALLENGES AND PROSPECTS

6.1 Challenges Faced

Although the DeepSeek large - scale model has achieved remarkable results, it still faces some challenges. In terms of model interpretability, due to the large number of model parameters and complex structure, it is difficult to understand the decision - making process of the model and the reasons for the output results.[6] When dealing with some sensitive information, how to ensure data security and privacy protection is also a problem that needs to be solved. The model's

performance depends on large - scale data and powerful computing resources, and limitations in data quality and computing resources may affect the model's effectiveness. In addition, the generalization ability in different fields and tasks and how to better adapt to complex and changeable practical application scenarios are also directions that require further research.

6.2 Future Development Directions

In the future, DeepSeek is expected to make breakthroughs in model interpretability research, developing visualization tools or interpretive algorithms to help users understand the internal mechanisms of the model. In terms of data security and privacy protection, more advanced encryption technologies and privacy - protection algorithms will be explored to ensure the security of data during use. With the development of hardware technology, more efficient computing devices and distributed computing technology will be utilized to further improve the model's training efficiency and performance. In applications, the model will be more deeply integrated into various industries, and more customized solutions will be developed to meet the needs of different users. There may also be explorations in cross - modal fusion, knowledge graph fusion, etc., to enhance the model's comprehensive capabilities.

7 CONCLUSION

The DeepSeek large - scale model, with its innovative technical architecture, unique training mechanism, and excellent performance, demonstrates strong competitiveness and broad application prospects in the field of artificial intelligence. It has achieved excellent results in tasks such as natural language processing, code generation, and multimodal understanding, providing strong support for the development of various industries. Although it faces some challenges, with the continuous progress of technology and in - depth research, the DeepSeek large - scale model is expected to make greater breakthroughs in the future, promoting the development of artificial intelligence technology and bringing more innovative applications and value to human society.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

8 REFERENCES

[1] Rohan Paul. DeepSeek - V3's Architectural Revolution: Rewriting the Economics of Large Language Model Training. 2024. Retrieved from https://rohanpaul.substack.com/p/deepseek-v3-technical-report-they

[2] Vaswani, A, Shazeer, N, Parmar, N, et al. Attention Is All You Need. Advances in Neural Information Processing Systems, 2017.

[3] DeepSeek-V3 Technical Report. It is authored by DeepSeek-AI, 2024. DOI: https://doi.org/10.48550/arXiv.2412.19437. Retrieved from https://arxiv.org/abs/2412.19437v1. Project homepage: https://github.com/deepseek-ai/DeepSeek-V3.

[4] Elmo. DeepSeek - VL: New Open Source Vision - Language Models! Medium (Medium Reviews). 2024. Retrieved from https://medium.com/@elmo92/deepseek-vl-new-open-source-vision-language-models-32bc77fa4647

[5] Xiaokang Chen, Zhiyu Wu, Xingchao Liu, et al. Janus-Pro: Unified Multimodal Understanding and Generation with Data and Model Scaling, 2025. Retrieved from https://arxiv.org/abs/2501.17811v1

[6] DeepSeek-AI, Daya Guo, Dejian Yang, et al. DeepSeek - R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. 2025. DOI: https://doi.org/10.48550/arXiv.2501.12948. Retrieved from https://arxiv.org/pdf/2501.12948

[7] DeepSeek-AI, Xiao Bi, Deli Chen, et al. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism. 2024. Retrieved from https://arxiv.org/abs/2401.02954