Proceedings of CONF-CDS 2025 Symposium: Application of Machine Learning in Engineering
DOI: 10.54254/2755-2721/2025.AST25999

Analysis of DeepSeek's Core Technology and its Effect on the Artificial Intelligence Area

Jiahua Zhang

Shanghai Shangde Experimental School, Shanghai, China
brucez113@outlook.com

Abstract. As recent years AI model developed, various new technology have been created, causing the AI model field to gradually become more competitive day after day, since it was born in 2023, DeepSeek gain benefit and was able to compete with other big companies using its benefits in code and mathematics advancement, localization and Chinese optimization, free to use, high performance. This thesis will analyze DeepSeek's core technology from its model architecture innovation, technology training and innovation, and performance comparison area. It will also explain DeepSeek's effect on the artificial intelligence area by focusing on DeepSeek's technology development in AI technology, the change in the competitive field of AI companies and ethical problems. In this thesis, we find that DeepSeek's ability to become a leader in the field of AI macromodels in just a few years is due to Deepseek's model architecture innovations, training strategy optimizations, and outstanding key performance, but like all other AI models, DeepSeek have limitations while having various benefits, these limitations include problems with the timeliness of data updates, as well as doubts about the accuracy and completeness of data and data biases.

Keywords: DeepSeek, core technology, AI model

1. Introduction

Artificial intelligence(AI) model is opening a new era, the development of foundation models is mainly separated into three periods: Germination period, Exploration and Precipitation period and the Rapid Development period, the Germination period lasts from 1950 to 2005, AI at the time was just a small amount of rule-based data processing, LeNet-5 was created during this period at 1998, the Precipitation period lasts from 2006 to 2019, at this period, AI models that is based on transformer architecture such as GAN(Generative Adversarial Network), Google Bert and GPT-2(Generative Pre-trained Transformer-2) developed by OpenAI have been created, the rapid development period started from 2020 until now, at this period AI models such as Turing-NLG, GPT3, MT-NLG(Megatron-Turing Natural Language Generation), Switch transformer, Pangu-Weather and M6 these foundation models have been created, DeepSeek was also created at this period, due to the rapid development of AI models, people's lives are widely affected [1]. This thesis will primarily focus on DeepSeek's innovation on model structures, strategy on training and innovation, performance comparison these areas to analyze DeepSeek's core technology characteristics, it will also explain DeepSeek's affect on artificial intelligence area by focusing on

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.AST25999

DeepSeek's technology development on the AI technology, the change in competitive field of AI companies and ethical problems. This thesis will present typical AI structure's development characteristics by analyzing DeepSeek's core technologies. It will also discuss possible problems for AI structures, in order to give basic knowledge to researchers new to the field and investigating problems in AI structures.

2. Background of AI models

AI models have different usages for different scenes and different business sectors, this causes the whole situation to become in free expressions, usually when speaking of AI's realm of usages, it can be separated into normal foundation models and business foundation models, the normal foundation models have the ability to generate contents, application to C-suites and to generate AI pictures, the business foundation models can be separated into industrial, medical and financial foundation models [2]. From the perspective of modes, AI models can be separated into single-mode models and multi-mode models, single-mode models are primarily used in areas such as natural language and visual computer code. The multi-mode models are primarily used in scientific calculations and other foundation model areas. Architecturally, the main algorithmic basis is the transformer as the mainstream.

DeepSeek has been a part of the most important AI models since it was created in 2023 due to its high performance, low cost and full open source traits. Its technology is positioned as general intelligence, open source ecology and cost revolution. In the general intelligence area, DeepSeek can be used in overlay text generation, code programming, mathematical processing and multi-mode processing. In the open source ecology area, DeepSeek supports full openness of model code and weights. It supports secondary development and business factors. Compared to most other AI models, it has relatively great advantages. At factors of cost, DeepSeek's cost of training is only one-twentieth the cost of the GPT. DeepSeek is mainly used in educational, financial quantification, healthcare and creative production areas [3].

3. Analysis of DeepSeek's core technology

AI model is based on the underlying framework. DeepSeek uses the transformer framework as its underlying framework. The transformer framework is a deep learning architecture based on a Multi-head Attention mechanism. It uses Self-Attention Mechanism to capture arbitrary positional relations in a sequence [4]. DeepSeek applied transformer structure's Multi-head Attention mechanism, Feed-Forward Networks and Residual Connections & Layer Normalization and innovated based on the basics of these technologies. This improves DeepSeek's runtime performance and efficiency.

3.1. Model architecture innovation

DeepSeek employs architectural innovations such as the Sparse Attention mechanism, the exploration of the MoE architecture (Mixture of Experts), optimizer and training strategies and improvements in positional encoding. This enhances DeepSeek's model running effectiveness and thinking speed [5-6].

DOI: 10.54254/2755-2721/2025.AST25999

3.1.1. Sparse attention mechanism

Sparse Attention mechanism reduces memory footprint and model processing time through a strategy of selecting specific locations for computation, these strategies includes localized windows, other sequences with the strongest relevance to the current content and modes gained from learning, from these strategies, DeepSeek can focus only on inputting relevant position in the sequence, thus reducing the cost for calculation and improves processing speed. Sparse Attention mechanism can decrease the complexities of calculation, reducing need of storage and improve long-distance dependent learning, Sparse Attention mechanism have significant advantage when processing long sequences, but it also have its own deficits: according to the variety of tasks, the programmers might have to use or develop different Sparse modes to complete the tasks, Sparse Attention mechanism might also ignore or miss some important information.

3.1.2. The exploration of the moe architecture

MoE architecture is based on Transformer architecture. It is made up by two sectors called sparse MoE level and door control network. Compared to other architectures that only use single networks, MoE can deploy separate network unit called “Experts”. these multi-network units allows MoE architecture to have advantages such as quick training speed, low computational costs, good scalability and able to learn multiple tasks at the same time, but because of the creation of multi-network units, MoE needs more resources to implement or modify, while communication is more costly.

3.1.3. Optimizer and training strategies

DeepSeek uses an optimizer that has the effect of an autonomous tuning mechanism helps the model to have faster running speed and decreases cost of running when being trained, Optimizer and training strategies have these advantages: Supports distributed training, breaks resource constraints, matches model architecture and maximizes performance. Optimizer is primarily used for decreasing the time used for outputting the correct option, avoiding the areas that could have caused damage and keeping the gradient stable [7].

DeepSeek combines optimizers such as AdamW and Lion with large-scale distributed training techniques such as parallel computing, gradient accumulation, and low-precision training in order to extract key information from massive amounts of data efficiently and consistently.

3.1.4. Improvements in positional encoding

Traditional Transformer coding uses fixed position coding, it will have deficits when running long sequences, DeepSeek improved and made new strategic improvements e on both RoPE (Rotary Position Embedding) and ALiBi (Attention with Linear Biases) position coding, separating the position coding into two parts increases running speed and decreases steps needed for each procedure, thus improves the availability of programs that rely on long distances and can process data from sequences better [8].

3.2. Training and optimization

DeepSeek has an excellent ability to find and answer programming and academic documents, and is even more intelligent than most other AIs, with good answers to most academic questions and even

DOI: 10.54254/2755-2721/2025.AST25999

code-related fields. Although DeepSeek has not invoked the image-drawing function, it is smarter and more accurate than other AIs in extracting information described by the user.

DeepSeek has not cited the image drawing function, but its user description of the information extraction is more intelligent and accurate than other AI. The reason why DeepSeek has such a powerful ability is mainly because its official team has made the following strategies: a. large-scale and diverse data sources; b. data cleaning and filtering; c. multi-language and multi-modal exploration.

Programmers add important data about web texts, books, academic papers, codes, conversations and answers to DeepSeek to enhance the quality of DeepSeek's answers. The web text adds a lot of frequently mentioned terms to DeepSeek. The books and academic papers provide DeepSeek with a lot of specialized knowledge and enhance DeepSeek's reasoning. The code base adds a lot of knowledge about the programming language, and the dialog and answer data allow DeepSeek to better engage in dialog with the user.

Since there is a lot of sensitive and even dangerous information on the Internet today, to prevent DeepSeek from adopting it, the team has made a number of strategies, such as removing repetitive text to prevent the system from overfilling a fixed viewpoint, filtering out low-quality data and information, minimizing discriminatory or stigmatizing data about different groups, and filtering out objectionable or inappropriate information.

Based on DeepSeek's response data, the team may have also used images, music, and other data to increase DeepSeek's comprehension of non-textual content and the accuracy of its responses when creating DeepSeek.

3.3. Comparison of core performances

The current most famous AI models in the field are DeepSeek, ChatGPT, ERNIE Bot, DoubaoAI, Qwen, etc. This thesis will analyze each AI model by focusing primarily on main specifics, application scenarios, core advantages, and arithmetic costs. The result is shown in Table 1.

Table 1. Comparison of different AI models [9]

Company name	DeepSeek	OpenAI	Baidu	ByteDance	Alibaba Corporation
AI model	DeepSeek	ChatGPT	ERNIE Bot	Doubao AI	Qwen
Main specifics	Ultra-long context support, multi-language modal understanding, freeware, powerful logical reasoning	Have excellent ability to understand language, suitable for scenes such as chatting on answering	Express clearly, supports multi-modal generation and processing of text, images, audio, etc.	Touching up the generated text for better user understanding	Because of its excellent logical thinking and communication ability, it is usually used for commercial purposes
Application scenarios	Programming, academic research and writing, logical interference, daily chat and creative generation	Generating chat, online education, content creation	Intelligent server, content suggestion, image and video analysis	Content creation, improve social media content	Large companies, economic service, business management
Core advantages	Code and mathematics advancement, localization and Chinese optimization, free to use, high performance	High-level language generation, multicast dialogue capability	Improve knowledge, multi-core processing	Natural language processing, generate and improve texts	Enterprise applications, supported by AliCloud

DOI: 10.54254/2755-2721/2025.AST25999

Table 1. (continued)

Arithmetic costs	DeepSeek-V3：<$0.01	GPT-4 Turbo: $0.01-$0.03	ERNIE：$0.002-$0.008	Doubao：$0.0015-$0.006	Qwen-72B：$0.002-$0.008
Trivial	Domestic Arithmetic Optimization	OpenAI uses high price strategy	Baidu SDK improvement	Presumably relies on self-research optimization + scale effect	Corresponding to DeepSeek-V3

4. DeepSeek's influence toward AI area

4.1. Technology improvement

DeepSeek improves the whole field's technology while improving itself: Open source ecology is where its codes are totally open to the public; this allows medium or small corporations or scientists to use advanced AI technology, no needing to write the code from the beginning. Architecture innovation is where its MoE and Sparse attention mechanism have been used by numerous developing teams and has become a new trend for the field. Training improvement is achieved with traditional methods. DeepSeek's training method allows foundational models to train faster and use less resources. Long-text processing is which it supports long texts, which directly increases the field's goals, and most companies are currently focusing on the problem of "AI short-term memory".

4.2. Industry competitive landscape

The birth of DeepSeek has made the commercial and technological competition in the field of AI more intense. In China, well-known companies such as Baidu and Alibaba have also had to adjust their strategies due to DeepSeek's outstanding performance in mathematics and programming. Because DeepSeek has some of its abilities opened freely to normal citizens, many international companies such as ChatGPT have to lower their prices. DeepSeek's advantage in specialized fields such as coding and academic research causes competitors to find new directions in order to have greater profits. DeepSeek is able to adapt to domestic chips and thus reduce the dependence on foreign hardware.

4.3. Ethical problems

DeepSeek's strong abilities allow more people to use its AI with less effort, but its specialization causes some people to become worried about DeepSeek. In order to solve these problems, DeepSeek used special mechanisms a. Content filtering is which DeepSeek have strict filtering mechanisms to prevent giving controversial content. b. Bias control is what DeepSeek has been more equalized in multi-language environments and decreases answers that might have caused stereotypes or discrimination. c. Risk of abuse is how to prevent the AI from being used by hackers. The team has been fortifying the AI network and applications. d. Transparency is which although DeepSeek opens its models to the public, its training data hasn't been fully opened to the public due to people worrying about "Blackmail" problems.

DOI: 10.54254/2755-2721/2025.AST25999

5. Future speculations and analysis

5.1. DeepSeek's limitations

DeepSeek is excellent, but far from perfect. a. The current system has low support for multi-modes. The current system is only for experts in processing texts; it cannot talk from pictures, such as GPT-4. b. Less adaptability toward unknown subjects. If the question is too abstract or obscure, DeepSeek's results will then be inaccurate. c. Energy-costly, although DeepSeek's creators advanced the model, it still needs a large amount of energy to operate the whole system. d. Like other AI models, DeepSeek also has problems with the timeliness of data updates, as well as doubts about the accuracy and completeness of data and data biases.

5.2. Future research areas

Afterwards, DeepSeek might develop from these directions: a. Smarter structures, in which DeepSeek could combine the Sparse Attention method and MoE close to each other in order to let calculation efficiency double. b. Synthetic data training is which using AI-generated data to help training, which decreases the reliance on realistic data. c. Lightweight deployment is where DeepSeek can be used on mobile applications such as phones; it doesn't need to rely on cloud anymore. d. Optimization of human-computer collaboration, which makes AI's explanations easier to understand, it allows humans to better understand and control them. DeepSeek can consider to have further research with other AI models, thus allows the cost to decrease, improves efficiency and gains access to more comprehensive and multifaceted data.

6. Conclusion

DeepSeek used its general AI, open-source ecology and cost revolution to have a foothold within this competitive AI model field within merely two years. By innovating the Transformer structure, exploring the MoE structure basis, and using autonomous adjustment mechanisms to optimize and train, this allows DeepSeek has specific advantages compared to most of the current AI models in the field. As DeepSeek develops, it reshapes the whole AI model field, pushes forward the advancement of AI technology, and changes the competitive goals of the AI field. Like other AI models, DeepSeek still has large amounts of problems that need to be solved, such as limitations of the technology and limitations of ethical problems. Thus, looking at the future, DeepSeek can gradually solve these existing problems of ethical and technological problems. Due to the limitation of space, this thesis does not analyze DeepSeek's technical route and the technical characteristics of other mainstream AI models for in-depth discussion. Therefore, in the future, people can analyze the technical characteristics of DeepSeek and other mainstream AI models in detail, as well as the technical parameters and application effects.

References

[1] CAI Rui, GE Jun, and SUN Zhe. Overview of the Development of AI Pre-trained Large Models [J/OL]. Journal of Chinese Mini-Micro Computer Systems, 2024: 1-12.

[2] Yufeng Wang. Unlocking a New Chapter in AI Large-Scale Model Applications: Technological Evolution, Challenges, and Future Prospects, 2024: 18-19

[3] Deng, Z., Ma, W., Han, Q. L., Zhou, W., Zhu, X., Wen, S., & Xiang, Y. Exploring DeepSeek: A Survey on Advances, Applications, Challenges and Future Directions. IEEE/CAA Journal of Automatica Sinica, 12(5), 2025: 872-893.

DOI: 10.54254/2755-2721/2025.AST25999

[4] Turner, R. E. An introduction to transformers. arXiv preprint arXiv: 2304.10557. 2023
[5] Martins, A., Farinhas, A., Treviso, M., Niculae, V., Aguiar, P., & Figueiredo, M. Sparse and continuous attention mechanisms. Advances in Neural Information Processing Systems, 33, 2020: 20989-21001.
[6] Masoudnia, S., & Ebrahimpour, R. Mixture of experts: a literature survey. Artificial Intelligence Review, 42, 2014: 275-293.
[7] Wang, C., & Kantarcioglu, M. A review of DeepSeek models' key innovative techniques. arXiv preprint arXiv: 2503.11486. 2025
[8] Gu, Z., Zhang, H., Chen, R., Hu, Y., & Zhang, H. Unpacking Positional Encoding in Transformers: A Spectral Analysis of Content-Position Coupling. arXiv preprint arXiv: 2505.13027. 2025
[9] Shi Zhenyu, Yu Haiyan, Zhang Kun, Liu Fangqi, Shen Dinglai, & Li Changbing. (2025).New Path for the Development of Management Science and Engineering Disciplines Integrating DeepSeek Large Models. Management Science and Engineering, 14, 640.