

Project Proposal Draft

Topic title: Deep Dive into Deepseek: Training Process, RL techniques, MoE, ...

Project title: Understanding DeepSeek: Architecture, Reasoning Capabilities and Training Processes

Authors: Hanna Azham, Marc Nauendorf and Nidhi Soni

1 Objectives

This project aims to provide an overview of DeepSeek’s model family and training strategies. The focus is on reinforcement learning techniques, the Mixture-of-Experts (MoE) architecture and system-level optimizations used across models like DeepSeek-R1, V2, Coder-V2, V3 and Math [1, 2, 3, 4, 5].

The first goal is to explain how DeepSeek fine-tunes its models. The analysis focuses on the training pipeline, especially how DeepSeek-R1 and R1-Zero apply reinforcement learning instead of traditional supervised learning [6, 7].

The second goal is to examine how DeepSeek combines reinforcement learning with chain-of-thought prompting. This helps to explain how reasoning becomes more accurate and logical in models like DeepSeek-Math [8].

The third goal is to study the role of the Mixture-of-Experts architecture. The project will show how MoE, used in DeepSeek-V2, V2.5 and V3, helps reduce computational costs while maintaining strong performance [9, 10].

The fourth goal is to investigate DeepSeek’s reward systems. The focus is on how these mechanisms improve output structure and logical reasoning, especially in DeepSeek-Math [2, 8].

The fifth goal is to explore the system-level improvements DeepSeek uses to optimize training and inference. This includes GPU optimizations such as kernel fusion and dynamic expert routing in large models like DeepSeek-V3 [11, 10].

The final goal is to compare DeepSeek models with other large language models like GPT-3.5 and LLaMA-2. This comparison will look at reasoning quality, efficiency and scalability [1].

2 Scope

This project focuses on the technical design and training process of the DeepSeek language models. It examines the full range of DeepSeek models, including DeepSeek-R1 for general reasoning, DeepSeek-V2 and V3 for advanced language understanding, DeepSeek-Coder-V2 for programming tasks and DeepSeek-Math for mathematical problem solving. Each model offers unique design choices such as reinforcement learning, Mixture-of-Experts (MoE) and reward-based training strategies. This project does not examine ethical, commercial or social issues. It also does not include historical or business aspects of DeepSeek’s development. The focus remains on the technical and architectural innovations across the DeepSeek model family.

3 Strategy

The project uses a research-based and analytical method to study DeepSeek’s training strategies and model architectures. The focus is on official papers, GitHub repositories, technical blogs and recorded explainers [5, 7]. Key scientific sources include peer-reviewed studies on DeepSeek-R1, DeepSeek-V2, DeepSeek-V3, Coder-V2 and DeepSeek-Math. These models cover areas such as reasoning, efficiency, coding and mathematical problem solving [1, 2, 12].

The project starts by outlining the development timeline of DeepSeek. This helps to understand changes from early models based on supervised fine-tuning to newer ones trained with reinforcement learning such as DeepSeek-R1-Zero. The next step focuses on how chain-of-thought prompting is used to improve reasoning in advanced models like DeepSeek-R1 and DeepSeek-V3. To understand the Mixture-of-Experts (MoE) approach used in DeepSeek-V2, V2.5, and V3, the project reviews model blueprints and routing logic. This helps explain how only task-relevant parts of the model are activated which reduces computing cost while maintaining quality [13, 4]. Reinforcement learning techniques are examined with a focus on Group Relative Policy Optimization (GRPO), as used in DeepSeek-Math. The project looks at how this reward system improves reasoning structure and output formatting. Reward definitions and training procedures are analyzed to support this part [2]. The project also examines GPU and system-level optimizations used in DeepSeek’s development. These include memory-efficient attention and dynamic expert routing which improve training and inference speed [11]. Benchmark results and task-specific examples from models like GPT-3.5 and LLaMA-2 are used for comparison. This provides a clear view of DeepSeek’s strengths in reasoning, efficiency, and scalability [1, 3].

4 Anticipated Outcomes

This project will give an overview of DeepSeek’s training methods and model structure. It will show how the model moves from pretraining to fine-tuning using reinforcement learning and chain-of-thought prompting. It will explain how models like DeepSeek-R1 use reinforcement learning to replace standard supervised fine-tuning. It will also show how DeepSeek-Math improves step-by-step reasoning using reward signals. The project will describe how Mixture-of-Experts architectures, used in DeepSeek-V2, V2.5 and V3, reduce computational costs by activating only the needed expert modules. It will examine GPU and system-level optimizations that improve speed and reduce resource use. The project will compare DeepSeek models with baseline models such as GPT-3.5 and LLaMA-2 by looking at reasoning ability, efficiency and scalability. The project will offer a helpful reference for researchers who want to understand or build similar systems. It will also help others understand how to build models that are both powerful and cost-effective [4, 5].

5 Submission Format

The final submission will be a written report in digital format. The report will follow a clear and structured layout. It will include an introduction, background information, research questions, methods, results, discussion and conclusion. Each section will focus on one main idea to help the reader follow the analysis easily. The report will also contain visual elements to support the explanations. These visual elements will make complex ideas easier to understand. This format is suitable because it presents information in a logical way. It allows for deep analysis while keeping the structure simple and clear. A written report is easy to share and read on different devices. It also supports the use of references which are important for academic work. The digital format makes it easy to include links to external sources such as research papers, videos and GitHub pages [7, 5]. This helps the reader explore the topic further if they want more details.

References

- [1] M. Tordjman, Z. Liu, M. Yuce, V. Fauveau, Y. Mei, J. Hadjadj, I. Bolger, H. Almansour, C. Horst, A. Parihar *et al.*, “Comparative benchmarking of the deepseek large language model on medical tasks and clinical reasoning,” *Nature Medicine*, 2025.
- [2] S. Kambhampati, K. Stechly, and K. Valmeekam, “(how) do reasoning models reason?” *Annals of the New York Academy of Sciences*, 2025.
- [3] S. Sandmann, S. Hegselmann, M. Fujarski, L. Bickmann, B. Wild, R. Eils, and J. Varghese, “Benchmark evaluation of deepseek large language models in clinical decision-making,” *Nature Medicine*, 2025.
- [4] R. Staff, “Deepseek rushes to launch new ai model as china goes all in,” <https://www.reuters.com/technology/artificial-intelligence/deepseek-rushes-launch-new-ai-model-china-goes-all-2025-02-25>, 2025, accessed: 2025-05-01.
- [5] D. AI, “Deepseek github organization,” <https://github.com/deepseek-ai>, 2025, accessed: 2025-05-01.
- [6] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi *et al.*, “Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning,” *arXiv preprint arXiv:2501.12948*, 2025.
- [7] A. Explained, “Deepseek-r1 paper explained - a new rl llms era in ai?” <https://www.youtube.com/watch?v=DCqQCLlIBUt=369s>, 2025, accessed: 2025-05-01.
- [8] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang, M. Zhang, Y. Li, Y. Wu *et al.*, “Deepseekmath: Pushing the limits of mathematical reasoning in open language models,” *arXiv preprint arXiv:2402.03300*, 2024.
- [9] A. Liu, B. Feng, B. Wang, B. Wang, B. Liu, C. Zhao, C. Deng, C. Ruan, D. Dai, D. Guo *et al.*, “Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model,” *arXiv preprint arXiv:2405.04434*, 2024.
- [10] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan *et al.*, “Deepseek-v3 technical report,” *arXiv preprint arXiv:2412.19437*, 2024.
- [11] J. Dreyer, “China made waves with deepseek, but its real ambition is ai-driven industrial innovation,” *Nature*, vol. 638, no. 8051, pp. 609–611, 2025.
- [12] R. Wang, J. He, and H. Liang, “Medicine’s j.a.r.v.i.s. moment: how deepseek-r1 transforms clinical practice,” *Journal of Thoracic Disease*, vol. 17, no. 3, pp. 1784–1787, 2025.
- [13] G. Conroy and S. Mallapaty, “How china created ai model deepseek and shocked the world,” *Nature*, vol. 638, no. 8050, pp. 300–301, 2025.