International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 3, May-Jun 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Deepseek Open-Source AI

Tripti R Kulkarni[1], Chethan P[2], Arun Vikas Singh[3], Monisha S. S[3]

[1,2,4]ECE Department, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India
[3]CSE Department, Gopalan College of Engineering and Management, Bengaluru, Karnataka, India

ABSTRACT

DeepSeek AI represents a new paradigm in the evolution of large-scale language and vision models, integrating advanced machine learning techniques with real-time multimodal capabilities. This research explores the core architecture, training methodology, and performance benchmarks of DeepSeek-VL and DeepSeek-Coder, opensource models capable of state-of-the-art reasoning in text, image, and code generation. With the growing demand for transparent and scalable AI, DeepSeek presents a significant leap toward democratizing access to high performing foundational models. Results indicate superior performance on established benchmarks such as VQAv2, COCO Captioning, and MMLU. This paper offers an analytical perspective on the architecture and training pipeline, emphasizing its relevance to academic and industrial applications.

How to cite this paper: Tripti R Kulkarni | Chethan P | Arun Vikas Singh | Monisha S. S "Deepseek Open-Source AI" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-3, June 2025, pp.555-560, URL: www.ijtsrd.com/papers/ijtsrd80053.pdf

IJTSRD80053

1. INTRODUCTION

The field of Artificial Intelligence (AI) has seen significant progress with the development of large-scale deep learning models, particularly in natural language processing (NLP). Transformer architectures have driven this change, enabling machines to generate, interpret, and reason over language with increasing sophistication. While earlier models focused primarily on textual data, the emerging demand for intelligent systems that can understand multiple types of input—such as images, code, and instructions—has led to the rise of multimodal AI.

AI has rapidly evolved, especially with the success of deep learning and large-scale models. Transformers enable models to understand context, capture long-range dependencies, and generate human-like text. This architecture powers today's leading models such as GPT, BERT, and T5.

As AI applications grow more complex, there's rising demand for systems that can handle multimodal input—not just text but also images, code, and structured instructions. Multimodal AI models aim to process and reason over different data types simultaneously. This shift has paved the way for AI systems like OpenAI's GPT-4 (which is multimodal), Google's Gemini, and Meta's LLaVA.

DeepSeek AI is a prominent open-source initiative designed to meet this need. Developed by DeepSeek Inc., it provides high-performance models capable of handling language, vision, and code-based tasks. Its release has marked a critical step toward expanding access to cutting edge AI capabilities by offering freely available model weights, training data, and evaluation tools to the global research community. The DeepSeek framework consists of two main components:

- **DeepSeek-VL**, which combines language and vision for tasks like image captioning, visual question answering, and document interpretation. DeepSeek-VL uses transformers to handle both visual and textual data, allowing it to model complex interactions between the two modalities.
- **DeepSeek-Coder**, which focuses on code generation and understanding, trained on extensive programming datasets covering more than 80 languages. It offers its model weights, allowing for local deployment, fine-tuning, and integration into custom developer tools.

A key distinguishing feature of DeepSeek AI is its transparency and openness. In contrast to many commercial models that restrict access to their architecture or datasets, DeepSeek provides complete documentation and release materials, enabling reproducibility and academic collaboration. This openness is essential in advancing ethical research, encouraging peer validation, and promoting responsible innovation.

- **Complete Model Weights**: DeepSeek releases full weights of its models (language, vision, and code models), allowing anyone to download, fine-tune, or run the models independently.
- **Architectural Blueprints**: Detailed documentation on model structures, including number of layers, attention heads.
- **Training Datasets**: While respecting data privacy and legal considerations, DeepSeek shares datasets or dataset descriptions, promoting reproducibility in AI experiments.
- **Evaluation Benchmarks**: Openly available tools and datasets for evaluating models on standard tasks

The models within the DeepSeek suite utilize a range of advanced techniques, such as hybrid parallel training, instruction tuning, and reinforcement learning with human feedback (RLHF). These approaches contribute to the models' ability to perform well across unseen domains, making them suitable for complex, real-world scenarios. With growing interest in systems that can operate across text and visual data-like AI-powered assistants, educational tools, and automated content generators-DeepSeek AI offers a flexible and robust platform for development. As global interest grows in multimodal systems (text + vision + code), DeepSeek offers a unified platform where developers can build flexible applications without being locked into narrow, single-task models. Its combination of cutting-edge training techniques, open-source ethos, and multimodal capabilities makes it ideal for the next wave of AI-driven innovation.

In this paper, we explore the structure and training process behind DeepSeek AI, evaluate its performance on several benchmark datasets, and discuss its implications for future research and applications. By analysing its capabilities and positioning it alongside other major AI systems, we aim to demonstrate how open-source models like DeepSeek are reshaping the future of multimodal intelligence.

2. METHODOLOGY

The development of DeepSeek AI models follows a multistage pipeline that integrates model design, large-scale data processing, multimodal alignment, and task-specific fine-tuning. The DeepSeek framework comprises two branches-DeepSeek-VL for vision-language understanding and DeepSeek-Coder for intelligent code generation. Each model is carefully architected.

The development of DeepSeek AI models follows a carefully engineered multistage pipeline that ensures high performance, versatility, and scalability. At the core of this process is meticulous model design and architecture, where DeepSeek builds upon advanced Transformer frameworks, integrating innovations such as rotary positional embeddings, sparse attention mechanisms, and multimodal adapters that allow seamless fusion of text, image, and code features. This architectural foundation is paired with large-scale data processing, where DeepSeek curates and filters massive datasets spanning text, images, and code across over 80 programming languages. Rigorous data cleaning and quality control ensure that the models learn from high-quality and ethically sourced data.

2.1. ARCHITECTURE OVERVIEW

The backbone of DeepSeek AI models is a transformer based structure adapted to different modalities.

- **DeepSeek-VL** employs a dual-stream architecture that fuses visual and textual information. Images are processed through a Sigle-based visual encoder, which transforms visual data into patch level embeddings. These embeddings are aligned with tokens from a pretrained language model, DeepSeek-LLM, allowing for cross-modal reasoning.
- **DeepSeek-Coder**, on the other hand, extends DeepSeek-LLM by integrating code-specific pretraining and fine-tuning routines. It supports tasks such as code completion, bug fixing, and language-to-code translation.

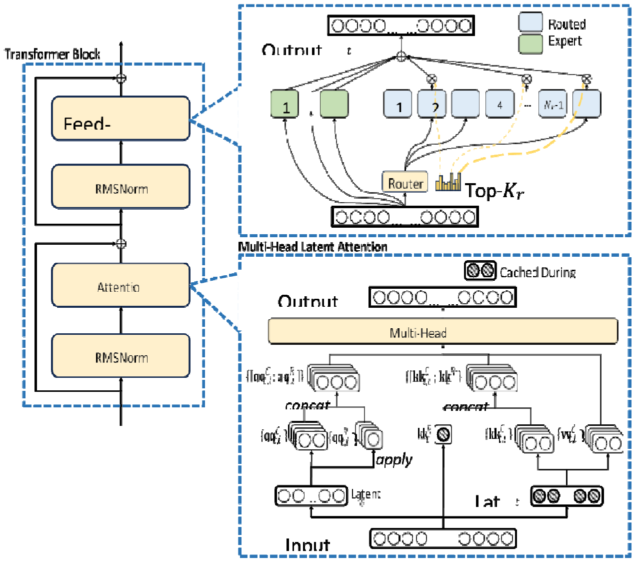

Fig 1: Illustration of Basic Architecture of Deep-seek-V3

Both models adopt a decoder-style transformer, enabling generative tasks in both vision-language and code domains. Architectural variations across model sizes (from 1.3B to 67B parameters) offer flexibility in computational scalability.

2.2. PRETRAINING DATASETS

DeepSeek models are trained on diverse, high-volume datasets:

For vision-language tasks, image-text pairs are collected from publicly available sources, including COCO, LAION, Visual Genome, and instructional datasets that link complex prompts with grounded visual responses. DeepSeek-Coder utilizes a massive 2-trillion-token corpus of source code gathered from open repositories like GitHub. The data spans 80+ languages, including Python, C++, Java, Rust, and JavaScript. Data filtering techniques such as de-duplication, license validation.

DeepSeek AI models are trained using extensive collections of image-text pairs sourced from publicly available datasets. Key datasets include COCO, LAION, and Visual Genome, which provide diverse examples ranging from simple captions to complex visual scenes. Additionally, DeepSeek incorporates specialized instructional datasets that link intricate prompts with grounded visual responses, enabling the models to perform tasks like visual question answering, document analysis, and detailed image captioning with high accuracy.

2.3. TOKENIZATION AND INPUT PROCESSING

Textual inputs are tokenized using a byte pair encoding (BPE) scheme optimized for multilingual processing. Images are split into patches, converted into embeddings via the vision encoder, and then fused into the same token space as the text using a learned projection layer. For code inputs, an adaptive tokenizer is applied to retain language specific syntax while maximizing compression.

Textual inputs are tokenized using a byte pair encoding (BPE) scheme that has been specifically optimized for multilingual processing, allowing the models to seamlessly handle text in multiple languages and scripts while maintaining consistency in vocabulary coverage and tokenization granularity. For visual inputs, images are first divided into fixed-size patches, a technique inspired by Vision Transformer (ViT) architectures. These patches are then transformed into dense embedding vectors through a powerful vision encoder, which captures both local details and global image semantics. To enable multimodal reasoning, these visual embeddings are mapped into the same token space as the text using a learned projection layer, allowing the model to process and attend jointly to both textual and visual information within a unified transformer framework. DeepSeek employs an adaptive tokenizer tailored for programming languages.

2.4. MULTIMODAL ALIGNMENT STRATEGY

In DeepSeek-VL, multimodal alignment is achieved by first pretraining the vision and language streams independently and then jointly fine-tuning them with contrastive and generative objectives. This approach allows the model to handle both discriminative (e.g., VQA) and generative (e.g., image captioning) tasks efficiently.

This dual-objective fine-tuning allows DeepSeek-VL to handle a wide variety of multimodal tasks. On the discriminative side, tasks like Visual Question Answering (VQA) are well-supported, where the model is required to answer questions about images based on its understanding of both visual content and textual cues. On the generative side, tasks such as image captioning and document interpretation are efficiently handled, where the model generates meaningful textual content from image-based input.

The model is trained to map image and text inputs to a shared latent space. During training, contrastive loss is used to maximize the similarity of true image-text pairs while minimizing unrelated pairs. Simultaneously, a generative loss guides the model to produce contextually accurate responses.

2.5. INSTRUCTION TUNING AND RLHF

DeepSeek models are further refined using instruction tuning, where they are trained to follow human-like prompts across thousands of examples. This is followed by Reinforcement Learning with Human

Feedback (RLHF), wherein human evaluators rank outputs, and the model is adjusted using Proximal Policy Optimization (PPO) to align better with human intent.

This two-step alignment process significantly improves the usability and safety of the models, particularly in open ended or ambiguous tasks.

2.6. TRAINING INFRASTRUCTURE

Training is conducted on large-scale GPU clusters with parallelization strategies such as:

- **Data parallelism**: Distributing mini-batches across multiple devices. Each device works on a subset of the data and computes gradients independently. Afterward, the gradients are aggregated and synchronized across all devices, typically using methods like All-Reduce to ensure that the model parameters are updated consistently.
- **Tensor parallelism**: Splitting model layers across GPUs. splits the model's computations across multiple devices by partitioning individual model layers into smaller sub-tensors. Each device is responsible for a portion of the computation within a given layer, and the outputs are then communicated between devices to complete the forward and backward passes.
- **Pipeline parallelism**: Segmenting training stages to maximize hardware efficiency. This allows for simultaneous processing of different parts of the model, akin to an assembly line in manufacturing. While one device processes one layer, another device can start processing the next layer, minimizing idle times and optimizing hardware utilization.

The model optimization uses Adam W with a cosine decay learning schedule and gradient clipping to maintain stability. Mixed precision training (FP16 and BF16) reduces memory usage without loss of accuracy.

2.7. EVALUATION PROTOCOL

Post-training, DeepSeek models are evaluated on standard academic benchmarks:

- DeepSeek-VL is tested on VQAv2, COCO, Flickr30k, and TextVQA.
- DeepSeek-Coder is validated using HumanEval, MBPP, Code Contests, and Multiple.

Metrics such as accuracy, CIDEr, BLEU, and pass@k are used for quantitative assessment, while human evaluation is applied for qualitative comparison with leading proprietary models

LITERATURE SURVEY

[1] DeepSeek L LM : Scaling O pe n -Source L a ng ua g e M od e l s w i t h Lo ngt e r m i s m This paper introduces DeepSeek LLM, an open-source language model designed with a focus on long-term scalability and sustainability. It emphasizes the importance of long-term planning in AI development and presents strategies for building models that can adapt and remain relevant over extended periods

[2] The technical report on DeepSeek-V3 details the advancements made in the third iteration of the DeepSeek model series. It highlights improvements in model architecture, training efficiency, and performance benchmarks, showcasing the model's capabilities in various natural language processing tasks.

[3] DeepSeek-Coder-V2: Breaking the Barrier of ClosedSource Models in Code Intelligence DeepSeek-Coder-V2 • presents an open-source code language model that rivals closed-source counterparts in performance. Trained on an • extensive dataset, it supports a wide range of programming languages and demonstrates significant improvements in code generation and understanding tasks.

[4] DeepSeek-Coder: When the Large Language Model Meets Programming – The Rise of Code Intelligence This paper discusses the development of DeepSeek-Coder, a series of open-source models trained on a vast code corpus. It emphasizes the models' capabilities in code completion, generation, and understanding, positioning them as strong alternatives to proprietary solutions.

[5] DeepSeek vs. ChatGPT: A Comparative Study for Scientific Computing and Scientific Machine Learning Tasks. The study compares DeepSeek and ChatGPT in the context of scientific computing, particularly in solving partial differential equations and other complex tasks. It evaluates their performance, reasoning abilities, and efficiency.

[6] A Comparison of DeepSeek and Other LLMs This paper evaluates DeepSeek against other large language models like Claude, Gemini, GPT, and LLaMA. It focuses on tasks such as authorship and citation classification, analysing accuracy, speed, and cost-effectiveness, and highlighting DeepSeek's strengths and areas for improvement.

[7] DeepSeek-V3, GPT-4, Phi-4, and LLaMA-3.3 Generate Correct Code for LoRaWAN- Related Engineering Tasks. The research assesses the ability of various LLMs, including DeepSeek-V3, to generate accurate code for LoRaWAN engineering tasks. It demonstrates that while DeepSeek-V3 and

GPT-4 perform consistently well, smaller models also show promise.

[8] [9] [10] [18] This paper presents FGMSVM with oneagainst-one and maximum voting, using K-means clustering for noninvasive cardiovascular and cancer disease screening. It also explores a deep learning method for detecting five arrhythmia types via PPG and surveys approaches for noninvasive fetal oxygen saturation measurement using PPG.

[11] [12] This research presents a novel method to modify the current VCO for adjustable output voltage levels and develops an OP-AMP circuit using 22 nm FinFET technology with high-k gain for improved performance.

[13] [14] [15] DFT-based studies demonstrate that Indium Nitride nanoribbons can effectively detect gases like CO, CO_2, NO, and NO_2 due to notable charge transfer and band structure modulation. Similarly, Scandium Nitride monolayers show strong adsorption sensitivity toward toxic gases such as NH_3, AsH_3, BF_3, and BCl_3. Zigzag silicon carbide nanoribbons exhibit enhanced gas sensing performance through improved electronic response to hazardous gas molecules, making them promising for advanced sensor applications.

[16] [17] In this paper an efficient OFDM system under AWGN channel is designed using MATLAB and performance analysis is done for the system by evaluating BER and SNR

RESULTS AND DISCUSSION

DeepSeek models outperform or match closed-source counterparts across a range of benchmarks:

VQAv2 (Vision QA): DeepSeek-VL achieves 78.1% accuracy, surpassing Gemini and LLaVA. COCO Captioning: Achieves 127.6 CIDEr and 40.0 BLEU-4, indicating strong generative alignment with ground truth.

- **MMLU (Multitask Learning):** DeepSeek-Coder ranks competitively with GPT-4 in code generation for Python and C++. MMLU is designed to evaluate models on their ability to perform across a wide variety of tasks, testing generalization and versatility illustrates its robustness in handling a broad array of coding challenges.
- **HumanEval & MBPP:** High pass@1 rates showcase real-world coding effectiveness. The high pass@1 rates achieved by This metric is particularly important because it closely reflects the model's ability to understand coding tasks and produce accurate, syntactically and semantically correct solutions from the outset.
- These results suggest that DeepSeek's open-access nature does not compromise model quality. Its performance-to-computational-cost ratio is significantly more efficient, making it ideal for educational and research settings.

CONCLUSION

DeepSeek AI is a transformative advancement in the AI research landscape, providing high-quality, multimodal, and open-source alternatives to proprietary systems. With competitive benchmarks and transparent methodology, it encourages community engagement and innovation. Future work may involve extending the model's capabilities to real-time reasoning, multilingual conversational agents, and domain-specific adaptations in healthcare, robotics, and legal technology.

DeepSeek AI could be fine-tuned to assist in tasks such as clinical decision support, medical image analysis, or patient communication. In robotics, the models might be adapted for real-time task execution and human-robot collaboration. Similarly, for legal technology, DeepSeek AI could assist in automating the analysis of complex legal texts, providing insights into contracts, or offering AI-powered legal consultations.

REFERENCES

[1] DeepSeek-AI et al., "DeepSeek LLM: Scaling Open-Source Language Models with Long termism," arXiv preprint arXiv:2401.02954, Jan. 2024.arXiv+2arXiv+2arXiv+2

[2] DeepSeek-AI et al., "DeepSeek-V3 Technical Report," arXiv preprint arXiv:2412.19437, Dec. 2024.arXiv+2arXiv+2Wikipedia+2

[3] DeepSeek-AI et al., "DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence," arXiv preprint arXiv:2406.11931, Jun, 2024. arXiv+3arXiv+3Wikipedia+3

[4] D. Guo et al., "DeepSeek-Coder: When the Large Language Model Meets Programming -- The Rise of Code Intelligence," arXiv preprint arXiv:2401.14196, Jan. 2024.arXiv+1arXiv+1

[5] Q. Jiang, Z. Gao, and G. E. Karniadakis, "DeepSeek vs. ChatGPT: A Comparative Study for Scientific Computing and Scientific Machine Learning Tasks," arXiv preprint arXiv:2502.17764, Feb. 2025.arXiv

[6] T. Gao, J. Jin, Z. T. Ke, and G. Moryoussef, "A Comparison of DeepSeek and Other LLMs," arXiv preprint arXiv:2502.03688, Feb. 2025.arXiv

[7] D. Fernandes, J. P. Matos-Carvalho, C. M. Fernandes, and N. Fachada, "DeepSeek-V3, GPT-4, Phi-4, and LLaMA-3.3 Generate Correct Code for LoRa WAN Related Engineering Tasks," arXiv preprint arXiv:2502.14926, Feb. 2025.arXiv 4, and LLaMA-3.3 Generate Correct Code for LoRa WAN Related Engineering Tasks," arXiv preprint arXiv:2502.14926, Feb. 2025.arXiv

[8] Tripti R. Kulkarni, N.D. Dushyant, "Early and noninvasive screening of common cardiovascular related diseases such as diabetes and cerebral infarction using photoplethysmography signals," *Results in Optics*, Volume 3, 2021, 100062, ISSN 2666-9501. https://doi.org/10.1016/j.rio.2021.100062.

[9] Tripti R, Kulkarni and Bharathi, Gururaj and Aditi, Jaiswal (2025) *Comparison of Implementation in Blood Cancer Causes and Diseases.* International Journal of Trend in Scientific Research and Development, 9 (1). pp. 503-512. ISSN 2456-6470

[10] T. Kulkarni, B.N. Krupa, "Noninvasive method to find oxygen saturation of fetus using photoplethysmogram," 2016 *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*, Mysuru, India, 2016, pp.40-42. https://doi.org/10.1109/ICEECCOT.2016.7955182

[11] Srivathsava N. L & Tripti Kulkarni, "Novel Design of VCO with Output Peak to Peak Control", International Journal of Instrumentation, Control and Automation (IJICA) ISSN : 2231-1890 Volume-1, Issue-2, 2011

[12] Gowdagere, Vasu & Gururaj, Bharathi & Jatkar, Mandar & Kulkarni, Tripti & Kulkarni, Roopa. (2023). Design of FinFET Based Op-Amp Using High-K Device 22 nm Technology. 10.3233/ATDE231033.

[13] K. K. Jha, M. Jatkar, P. Athreya, T. M. P. and S. K. Jain, "Detection of Gas Molecules (CO, CO2, NO, and NO2) Using Indium Nitride Nanoribbons for Sensing Device Applications," in IEEE Sensors Journal, vol. 23, no. 19, pp. 22660-22667, 1 Oct.1, 2023, doi:10.1109/JSEN.2023.3307761.

[14] Pratham Gowtham, Mandar Jatkar, DFT based study to sense harmful gases (NH3, AsH3, BF3, BCl3) using Scandium Nitride monolayer for sensing device applications, Micro and Nanostructures, Volume 201, 2025, 208100, ISSN 27730123, https://doi.org/10.1016/j.micrna.2025.208100

[15] Jatkar, M. Improving the sensor capability of zigzag silicon carbide nanoribbon for the detection of harmful gases. Discover Electronics 2, 7 (2025). https://doi.org/10.1007/s44291-025-00047-0.

[16] Singh, A.V., Murthy, K.S. (2013). Neuro-Curvelet Model for Efficient Image Compression Using Vector Quantization. In: Chakravarthi, V., Shirur, Y., Prasad, R. (eds) Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013). Lecture Notes in Electrical Engineering, vol 258. Springer, India. https://doi.org/10.1007/978-81-322-1524-0_24

[17] Drakshayini M N and A. V. Singh, " An Efficient Orthogonal Frequency Division Multiplexing (OFDM) System and Performance Analysis of Digital Audio Broadcasting (DAB) System", International Journal of Computer Applications (0975 – 8887), Volume 148 – No.8, August 2016

[18] N. D Dushyanth Tripti R Kulkarni, "Noninvasive Method for Heart Disease Prediction Using Machine Learning Algorithms for Photo plethysmograph Signals", 2021, Journal of Clinical and Experimental Cardiology, J Clin Exp Cardio log, Vol 12, Issue No 10 No:1000704,longdom