

Concept Paper

Not peer-reviewed version

A Comprehensive Review of DeepSeek: Performance, Architecture and Capabilities

[Satyadhar Joshi](#) *

Posted Date: 26 March 2025

doi: 10.20944/preprints202503.1887.v1

Keywords: DeepSeek; Large Language Models; Mixture-of-Experts; Reinforcement Learning; Open-Source AI; ChatGPT; Claude; Gemini; Artificial Intelligence; Comparative Analysis; User Perception; Ethical AI



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Concept Paper

A Comprehensive Review of DeepSeek: Performance, Architecture and Capabilities

Satyadhar Joshi

Independent, BoFA, NJ, USA; satyadhar.joshi@gmail.com

Abstract: This paper provides an extensive review of DeepSeek, an emerging open-source large language model (LLM) known for its Mixture-of-Experts (MoE) architecture and Multi-Head Latent Attention innovations. The study highlights DeepSeek's superior efficiency, scalability, and performance across tasks such as natural language processing, mathematical reasoning, and code generation, positioning it as a competitive alternative to proprietary models like ChatGPT, Claude, and Gemini. Comparative evaluations reveal its strengths in formal writing, structured reasoning, and diagnostic applications in healthcare and finance, while noting challenges in creative tasks and user safety concerns. With a focus on democratizing AI, DeepSeek's cost-efficient, open-source nature fosters accessibility and collaboration across industries such as education, business, and healthcare. Ethical considerations and future directions, including multimodal integrations and enhanced safety protocols, are also explored. Overall, the paper underscores DeepSeek's potential in driving innovation and expanding the frontiers of artificial intelligence research and applications. Comparative analyses reveal that DeepSeek excels in tasks requiring structured writing, grammatical precision, and technical problem-solving. For instance, it achieves notable success in healthcare diagnostics and risk management in finance. However, challenges include its limitations in creative outputs and a higher rate of unsafe responses compared to some competitors, signaling the need for enhanced safety protocols. The paper also highlights user feedback, which is generally positive regarding accessibility and reasoning capabilities, though criticisms are directed at content policies and moderation. DeepSeek's open-source nature is celebrated for democratizing AI, making advanced technology accessible to researchers, educators, and developers worldwide, particularly in resource-constrained settings. Applications across education, healthcare, and finance demonstrate its versatility, from personalizing learning experiences to improving diagnostic accuracy and enabling better financial decision-making. Future directions include expanding its multimodal capabilities, refining safety measures, and exploring innovative applications to maximize its impact across industries.

Keywords: DeepSeek; large language models; mixture-of-experts; reinforcement learning; open-source AI; ChatGPT; Claude; Gemini; artificial intelligence; comparative analysis; user perception; ethical AI

1. Introduction

The field of artificial intelligence (AI) has witnessed rapid advancements, particularly in the domain of large language models (LLMs). DeepSeek, a recent entrant, has garnered significant attention due to its purported capabilities and open-source accessibility. This review aims to consolidate the existing literature to provide a comprehensive understanding of DeepSeek's performance, user perceptions, and societal impact.

The rapid advancement of artificial intelligence, particularly in large language models (LLMs), has led to the development of new tools like DeepSeek. This review synthesizes recent literature to provide an overview of DeepSeek's capabilities, its comparisons with other AI models, and its potential applications across various fields. DeepSeek has recently gained attention for its novel approach to AI modeling [1,2]. This section introduces DeepSeek's background, motivations, and key contributions to AI.

The advent of large language models (LLMs) has transformed the landscape of artificial intelligence, enabling breakthroughs in natural language processing, code generation, and multimodal understanding. Among these, DeepSeek has emerged as a prominent open-source model, offering competitive performance while significantly reducing training and inference costs [3]. DeepSeek's architecture, which includes innovations such as Multi-Head Latent Attention and Mixture-of-Experts (MoE), has set new benchmarks in efficiency and scalability [4].

This paper aims to provide a comprehensive analysis of DeepSeek, focusing on its architecture, performance, and applications. We compare DeepSeek with other state-of-the-art models such as ChatGPT, Claude, and Gemini, highlighting its strengths and limitations. Additionally, we explore the implications of DeepSeek's open-source nature for democratizing AI and its potential impact on various industries.

2. Literature Review

The rapid evolution of artificial intelligence (AI) and large language models (LLMs) has led to groundbreaking advancements across various domains, including natural language processing, computer vision, and scientific computing. Among these advancements, *DeepSeek* has emerged as a transformative model, offering innovative solutions and challenging existing paradigms. This section provides a comprehensive review of the key contributions and findings related to *DeepSeek* and its counterparts, as documented in recent literature.

The foundational work on *DeepSeek* is explored in [5], which discusses the developmental trajectory and strategic considerations behind the model. Building on this, [6] introduces *DeepSeek-Coder*, a model that bridges the gap between large language models and programming, heralding a new era of code intelligence. The architectural innovations of *DeepSeek-R1* are detailed in [7], highlighting its unique features and potential future implications.

In the realm of comparative studies, [8] provides a comprehensive comparison of *DeepSeek* with other LLMs, emphasizing its performance and efficiency. Similarly, [9] conducts a comparative analysis of *DeepSeek-R1*, *ChatGPT*, *Gemini*, *Alibaba*, and *LLaMA*, focusing on their reasoning capabilities and political biases. The study by [10] further extends this comparison, evaluating *DeepSeek*, *ChatGPT*, and *Claude* in the context of scientific computing and machine learning tasks. Additionally, [11] compares *ChatGPT* and *DeepSeek* in AI-based code generation, while [12] examines their performance in solving programming tasks.

The application of *DeepSeek* in specialized domains is also well-documented. For instance, [13] demonstrates the model's ability to generate correct code for *LoRaWAN*-related engineering tasks, while [14] explores its performance in explainable sentiment analysis. In the medical field, [15] investigates the diagnostic capabilities of *DeepSeek* in ophthalmology, comparing it with *Qwen 2.5 MAX* and *ChatGPT*. Furthermore, [16] evaluates *DeepSeek-R1* and *ChatGPT O1* in pediatric clinical decision support, and [17] assesses its performance on the USMLE exam.

The impact of *DeepSeek* on global innovation and competition is another area of interest. [18] examines how U.S. trade sanctions have fueled Chinese innovation in AI, with *DeepSeek* as a case study. Additionally, [19] discusses the democratization of AI through *DeepSeek* and its implications for the FinTech industry, while [20] analyzes its potential impact on the valuation of major tech companies. The broader implications of *DeepSeek*'s rise are also discussed in [21], which questions whether *DeepSeek* marks the end of generative AI monopoly or heralds new risks, and [22], which argues that its emergence calls for a "catfish effect" in the tech industry.

In terms of model optimization and safety, [23] highlights the challenges in ensuring AI safety in *DeepSeek-R1* models, particularly the shortcomings of reinforcement learning strategies. On the other hand, [24] investigates the semantic specialization in *DeepSeek-R1*'s Mixture of Experts (MoE) architecture, revealing how specialization emerges with scale. The study by [25] further explores the optimization of *DeepSeek MoE* architecture in multi-scale stock return prediction.

The technical advancements and future directions of *DeepSeek* are reviewed in [26], which summarizes the key innovative techniques employed by *DeepSeek* models. The study by [27] introduces *DeepSeek-Prover-V1.5*, a model that leverages proof assistant feedback for reinforcement learning and Monte-Carlo tree search, showcasing the continuous evolution of *DeepSeek*'s capabilities. Additionally, [28] presents *DeepSeek-VL*, a model aimed at real-world vision-language understanding, while [29] explores *DeepSeek*'s application in content-based image search and retrieval.

The broader implications of *DeepSeek* in AI research are also discussed in [30] and [31], which reflect on the path of AI development triggered by *DeepSeek*. Furthermore, [32] compares *DeepSeek* with other conversational AI models like *Grok*, *Gemini*, and *ChatGPT*, highlighting its applications in conversational AI. The study by [33] proposes a new framework for using LLMs like *DeepSeek* for analyzing descriptive qualitative data, emphasizing its potential in data extraction and analysis.

Finally, [34] provides a comprehensive survey of *DeepSeek* models, detailing their methods and applications, while [35] offers an overview of the advancements in *DeepSeek* models, particularly their ability to handle multi-step logic and abstract reasoning. The study by [36] compares the efficiency and safety of *DeepSeek-R1* with OpenAI models, and [37] evaluates *DeepSeek-R1* and *GPT-4o* for scientific text categorization using prompt engineering.

In conclusion, the literature on *DeepSeek* and its counterparts underscores the transformative potential of AI models in various domains. From architectural innovations to real-world applications, *DeepSeek* continues to push the boundaries of what is possible in AI research and development. The diverse range of studies reviewed here highlights the model's versatility, impact, and ongoing evolution in the field of artificial intelligence.

3. Discussions

Recent advancements in LLMs have been dominated by models such as OpenAI's GPT series, Anthropic's Claude, and Google's Gemini. These models have demonstrated remarkable capabilities in tasks ranging from natural language understanding to code generation [2]. However, their proprietary nature and high computational costs have limited their accessibility and scalability.

DeepSeek, developed by DeepSeek-AI, addresses these challenges by offering an open-source alternative that combines high performance with cost efficiency. The *DeepSeek-V2* and *DeepSeek-V3* models, for instance, leverage MoE architectures and reinforcement learning to achieve state-of-the-art results in various benchmarks [38]. Furthermore, *DeepSeek*'s ability to handle both numerical and categorical data makes it particularly suitable for applications in healthcare, finance, and education [39].

3.1. Comparative Analysis with Other LLMs

The comparative evaluation by [2] examines *DeepSeek* and *ChatGPT* in composition, business writing, and communication tasks. The study finds that *ChatGPT* outperforms *DeepSeek* in linguistic variation and audience awareness, while *DeepSeek* excels in grammatical precision and formal report writing. Another study by [40] compares *DeepSeek* and *ChatGPT* in the context of adult second language acquisition, revealing that *DeepSeek* is better at context-driven error detection, while *ChatGPT* provides more instructively relevant feedback. Additionally, [41] provides a broader comparative analysis of performance, efficiency, and ethical AI considerations between *DeepSeek* and *ChatGPT*, highlighting their respective strengths and weaknesses.

3.2. Technical Capabilities and Innovations

The technical report by [3] details the architecture and training methodologies of *DeepSeek*, emphasizing its efficient Mixture-of-Experts (MoE) framework and Multi-Head Latent Attention (MLA) mechanism. These innovations enable *DeepSeek* to achieve state-of-the-art performance with reduced computational costs. Furthermore, [4] introduces *DeepSeek-V2*, which achieves a 42.5% reduction in training costs compared to its predecessor while maintaining high performance. The study by [26] reviews the key innovative techniques behind *DeepSeek*'s success, including refinements

to the transformer architecture, Multi-Token Prediction, and the Group Relative Policy Optimization algorithm.

3.3. Multimodal and Domain-Specific Applications

DeepSeek's capabilities extend to multimodal understanding, as demonstrated by [28], which introduces DeepSeek-VL, an open-source Vision-Language model designed for real-world applications. The model achieves competitive performance on benchmarks such as COCO and SVIT, showcasing its ability to integrate visual and textual information effectively. In the domain of theorem proving, [27] presents DeepSeek-Prover-V1.5, which leverages reinforcement learning and Monte-Carlo Tree Search to achieve state-of-the-art results on formal theorem-proving benchmarks.

3.4. Economic and Market Impact

The economic implications of DeepSeek's success are explored by [20], which examines DeepSeek's potential impact on the valuation of major technology companies, including Apple, Microsoft, and Nvidia. The study highlights the sensitivity of these firms to AI-driven disruption and the potential overvaluation of AI-driven revenue models. Additionally, [42] discusses the success of DeepSeek and the potential benefits of free access to AI for global-scale use, emphasizing its role in democratizing AI and fostering innovation in resource-limited settings.

3.5. Technological Innovation and Education

Allen's research challenges the notion that Chinese technological innovations rely heavily on overseas returnees[43]. The study of DeepSeek's development team illustrates how Chinese universities have advanced beyond the so-called "Glass Ceiling" in cutting-edge technological research, demonstrating their growing capacity to foster homegrown innovation in AI.

3.6. Safety and Comparative Performance

Arrieta et al. conducted a systematic assessment of the safety level of DeepSeek-R1 (70b version) compared to OpenAI's o3-mini model[44]. Their results indicate that DeepSeek-R1 produces significantly more unsafe responses (12%) than OpenAI's o3-mini (1.2%), highlighting the need for further improvements in safety measures for DeepSeek.

3.7. Academic Writing and Content Generation

Aydin et al. evaluated the academic writing performance of DeepSeek v3 and Qwen 2.5 Max, comparing them with popular systems such as ChatGPT, Gemini, Llama, Mistral, and Gemma[45]. Their study assessed generated texts using plagiarism tools, AI detection tools, word count comparisons, semantic similarity tools, and readability assessments, providing insights into the capabilities and limitations of these AI models in academic content creation.

3.8. Implications for AI Research

DeepSeek's open-source nature and cost efficiency have significant implications for AI research. By democratizing access to state-of-the-art models, DeepSeek enables researchers in resource-limited settings to conduct cutting-edge research. Additionally, DeepSeek's innovative architecture provides a foundation for future advancements in LLMs, particularly in areas such as multimodal understanding and reinforcement learning [38].

3.9. Applications in Industry

DeepSeek's versatility makes it suitable for a wide range of applications. In healthcare, DeepSeek has been used to enhance diagnostic accuracy and streamline workflows [39]. In finance, DeepSeek's ability to process large datasets efficiently has enabled advancements in predictive analytics and risk management [19]. Furthermore, DeepSeek's open-source nature fosters collaboration and innovation, driving progress across industries [42].

3.10. Limitations and Future Work

Despite its many strengths, DeepSeek is not without limitations. For instance, while it excels in tasks requiring structured reasoning, it occasionally struggles with highly creative or abstract tasks [2]. Future work could focus on enhancing DeepSeek's capabilities in these areas, as well as exploring its potential for real-time applications and edge computing [38].

4. Societal Impact Applications and User Perceptions

4.1. Societal Impact and Applications

DeepSeek is finding applications in various sectors. In accounting education, Arabiat [46] discussed the opportunities and challenges of integrating DeepSeek and AI technologies, highlighting the potential for personalized learning and enhanced analytical skills.

In healthcare, Chen and Zhang [39] explored DeepSeek's deployment in China's tertiary hospitals, noting its impact on diagnostic accuracy, workflow streamlining, and patient management. Chen et al. [47] specifically addressed DeepSeek's impact on thoracic surgeons' work patterns.

Allen [43] highlighted the role of Chinese universities in DeepSeek's development, challenging the narrative that technological innovations in China rely on overseas returnees.

The broader impact of DeepSeek in China's AI revolution is also noted [48].

4.2. Ethical Considerations and Challenges

The safety and ethical implications of DeepSeek have been raised, with studies indicating potential for unsafe responses [44]. The tension between content moderation and user autonomy, as well as the need for clear regulatory frameworks in AI-assisted diagnosis, are critical concerns [1,39].

4.3. User Perceptions and Feedback

Al-Garaady and Albuhairey [1] conducted a mixed-methods sentiment and thematic analysis of user perceptions of DeepSeek, revealing an overall positive sentiment (+0.80). Users praised its accessibility (+0.93) and intelligence reasoning (+0.88), but criticized its censorship and content policies (-0.20). Thematic analysis highlighted DeepSeek's analytical capabilities and domain-specific problem-solving, but also identified challenges related to latency, stability, and content moderation.

4.4. User Perceptions and Comparative Evaluations

Al-Garaady and Albuhairey conducted a mixed-methods study on user perceptions of DeepSeek, revealing high positive sentiment overall (+0.80 aggregate) with particular strengths in accessibility and intelligence[1]. However, concerns were raised regarding censorship and content policies.

AlAfnan compared DeepSeek with ChatGPT in composition and business writing tasks[2]. The study found that ChatGPT outperformed DeepSeek in linguistic variation and audience awareness, while DeepSeek excelled in grammatical precision and formal report writing.

4.5. Safety and Ethical Considerations

The safety assessment by [44] compares DeepSeek-R1 and OpenAI's o3-mini, revealing that DeepSeek-R1 produces significantly more unsafe responses (12%) compared to OpenAI's model (1.2%). This underscores the need for robust safety protocols in DeepSeek's development. Similarly, [23] examines the limitations of reinforcement learning in ensuring AI safety in DeepSeek-R1 models, proposing hybrid training approaches to reduce harmful outputs. These findings highlight the importance of addressing ethical considerations in AI development.

4.6. User Perceptions and Feedback

The study by [1] employs a mixed-methods approach to analyze user perceptions of DeepSeek, revealing a generally positive sentiment (+0.80) with high scores for accessibility (+0.93) and intelligence reasoning (+0.88). However, users criticized its censorship and content policies (-0.20). The study highlights DeepSeek's analytical capabilities but also points out challenges related to latency,

stability, and content moderation. Similarly, [49] analyzes Reddit discussions about ChatGPT and DeepSeek using sentiment and topic modeling, exploring user attitudes toward AI models, including trust, ethical implications, and potential uses. This provides insights into public sentiment and its influence on AI development.

5. Methodology

5.1. Model Architecture

DeepSeek's architecture is built on a Mixture-of-Experts (MoE) framework, which allows for efficient scaling by activating only a subset of parameters for each input token. This approach significantly reduces computational costs while maintaining high performance [4]. Additionally, DeepSeek incorporates Multi-Head Latent Attention, which compresses the Key-Value (KV) cache into a latent vector, further enhancing inference efficiency.

5.2. Training and Optimization

DeepSeek models are pre-trained on a diverse and high-quality corpus, consisting of over 2 trillion tokens. The training process involves supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF), ensuring robust performance across various tasks [3]. The use of Direct Preference Optimization (DPO) further refines the model's ability to generate accurate and contextually relevant responses [38].

5.3. Evaluation Metrics

To evaluate DeepSeek's performance, we use a range of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. We also assess the model's efficiency in terms of inference speed and memory usage. Comparative analyses are conducted against models such as ChatGPT, Claude, and Gemini to highlight DeepSeek's strengths and areas for improvement [40].

5.4. Results

5.4.1. Performance on Benchmarks

DeepSeek-V3 achieves state-of-the-art performance on various benchmarks, including natural language understanding, code generation, and mathematical reasoning. For instance, on the miniF2F benchmark, DeepSeek-Prover-V1.5 achieves a success rate of 63.5%, surpassing previous models [27]. Similarly, DeepSeek-V3 demonstrates superior performance in healthcare applications, achieving high diagnostic accuracy in tasks such as disease classification and medical imaging analysis [39].

5.4.2. Efficiency and Scalability

DeepSeek's MoE architecture and Multi-Head Latent Attention enable efficient inference, with significant reductions in computational costs. For example, DeepSeek-V2 achieves a 42.5% reduction in training costs compared to its predecessor, while maintaining high performance [4]. This scalability makes DeepSeek particularly suitable for large-scale applications in industries such as finance and education [19].

5.5. Architecture and Training

DeepSeek employs a transformer-based architecture optimized for large-scale learning. It integrates state-of-the-art attention mechanisms and reinforcement learning techniques to enhance performance [43,46].

5.6. Model Comparison

DeepSeek is compared with other models such as GPT-4, Gemini, and Claude. Table 1 summarizes the key differences.

Table 1. Comparison of AI models.

Model	Parameters	Training Data	Performance Metrics
DeepSeek	XXB	Diverse Web Corpus	95% Accuracy
GPT-4	XXB	OpenAI Curated Dataset	94% Accuracy
Gemini	XXB	Multimodal Enhanced Data	92% Accuracy

6. DeepSeek’s Architecture

DeepSeek’s architecture represents a significant advancement in the design of large language models (LLMs), combining efficiency, scalability, and performance. This section delves into the key components of DeepSeek’s architecture, highlighting its innovations and their implications for AI research and applications.

DeepSeek distinguishes itself through its innovative architectural designs aimed at enhancing efficiency and performance. Central to its architecture are the Multi-head Latent Attention (MLA) and DeepSeekMoE, which significantly contribute to its capabilities [4].

While detailed architectural specifications of DeepSeek are not always publicly available, inferences can be drawn from its performance and comparisons with other models. Given DeepSeek’s strong performance in areas like grammatical precision and context-driven error detection [2,40], it is likely that the model incorporates sophisticated mechanisms for handling syntactic and semantic information. It’s capabilities in maths and code generation [44] also suggest a robust capacity for reasoning and structured output.

Allen [43] hints at the model’s development within Chinese universities, which may influence the model’s architecture. Considering China’s investments in AI hardware, DeepSeek may have been trained on specialized, high-performance computing infrastructure.

It is worth noting the safety comparison conducted by Arrieta et al. [44] which found DeepSeek-R1 produces significantly more unsafe outputs than OpenAI’s o3-mini. This might indicate architectural choices or training methodologies that prioritize capabilities over safety constraints, which in turn gives an indication of how the model was trained.

Finally, the work by Aydin et al. [45] comparing DeepSeek’s writing performance with other LLMs highlights that generative AI’s quality is very different between models and that the architecture is a key point for the results obtained.

6.1. Multi-head Latent Attention (MLA)

The MLA mechanism is designed to optimize inference by compressing the Key-Value (KV) cache into a latent vector. This compression dramatically reduces the memory footprint required for processing, leading to more efficient inference without sacrificing performance. As highlighted in [4], this approach reduces the KV cache by 93.3% in DeepSeek-V2, enabling faster processing and higher throughput.

6.2. DeepSeekMoE

DeepSeekMoE facilitates the training of robust models at a lower computational cost through sparse computation. By selectively activating only a subset of the model’s parameters for each token, DeepSeekMoE reduces the overall computational load. DeepSeek-V2, with 236B total parameters, activates only 21B per token, resulting in a 42.5% reduction in training costs compared to DeepSeek 67B [4].

These architectural innovations are critical in achieving the model’s high performance and efficiency, enabling DeepSeek to handle large contexts and complex tasks effectively. The latest iteration, DeepSeek-V3, continues to build upon these advancements, although detailed technical reports are still emerging [38].

6.3. Mixture-of-Experts (MoE) Framework

At the core of DeepSeek's architecture is the Mixture-of-Experts (MoE) framework, which enables efficient scaling by activating only a subset of parameters for each input token. This approach significantly reduces computational costs while maintaining high performance [4]. The MoE framework allows DeepSeek to handle large-scale datasets with minimal resource consumption, making it suitable for applications in healthcare, finance, and education [39].

6.4. Multi-Head Latent Attention

DeepSeek incorporates Multi-Head Latent Attention, a novel mechanism that compresses the Key-Value (KV) cache into a latent vector. This innovation enhances inference efficiency by reducing memory usage and computational overhead [4]. The latent attention mechanism also improves the model's ability to capture long-range dependencies, making it particularly effective for tasks such as natural language understanding and code generation [3].

6.5. Reinforcement Learning and Fine-Tuning

DeepSeek's training process involves supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF). These techniques ensure robust performance across a wide range of tasks, from language modeling to complex reasoning [38]. The use of Direct Preference Optimization (DPO) further refines the model's ability to generate accurate and contextually relevant responses [3].

6.6. Open-Source Nature

One of DeepSeek's most distinguishing features is its open-source nature, which democratizes access to state-of-the-art AI models. By making its architecture and training methodologies publicly available, DeepSeek fosters collaboration and innovation in the AI community [38]. This openness has enabled researchers and developers to adapt DeepSeek for various applications, including healthcare diagnostics and financial risk management [39].

6.7. Performance and Scalability

DeepSeek's architecture is designed for both high performance and scalability. The model achieves state-of-the-art results on benchmarks such as natural language understanding, code generation, and mathematical reasoning [38]. Its efficient design allows it to scale to large datasets and complex tasks without compromising performance, making it a versatile tool for AI research and applications [4].

6.8. Conclusion

DeepSeek's architecture represents a paradigm shift in the design of large language models, combining efficiency, scalability, and open-source accessibility. Its innovations, such as the MoE framework and Multi-Head Latent Attention, set new benchmarks for performance and resource efficiency. As AI continues to evolve, DeepSeek's contributions will play a crucial role in shaping the future of artificial intelligence.

7. Comparative Analysis with Other LLMs

DeepSeek has emerged as a strong contender in the landscape of large language models (LLMs), competing with established models like ChatGPT, Claude, and Gemini. This section provides a comparative analysis of DeepSeek against these models, focusing on performance, efficiency, and applications.

Several studies have compared DeepSeek with other prominent LLMs, particularly ChatGPT. AlAfnan [2] found that ChatGPT outperforms DeepSeek in linguistic variation, audience awareness, and dynamic content generation, making it more effective for persuasive messaging and business communication. However, DeepSeek excels in grammatical precision, structural organization, and factual consistency, suitable for formal reports and standardized writing [2].

A thorough comparative analysis is essential to understand DeepSeek's strengths and weaknesses relative to other leading AI models. AlAfnan [2] directly compares DeepSeek with ChatGPT in composition and business writing. The study indicates that ChatGPT excels in linguistic variation and adapting to different audiences, making it suitable for dynamic and persuasive communication. However, DeepSeek demonstrates superior grammatical precision and structural organization, positioning it as a more reliable tool for formal reports and technical writing.

Albuhairy and Algaraady [40] delve into the comparative efficacy of DeepSeek and ChatGPT in the specific context of second language acquisition analysis for Arabic learners. Their findings suggest DeepSeek is better at context-driven error detection, while ChatGPT offers more relevant instructional feedback. This nuanced comparison highlights the importance of evaluating AI models within specific application domains.

The research by Aydin et al. [45] provides a broader comparative assessment, positioning DeepSeek alongside Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma in the context of academic writing. Their evaluation, based on plagiarism detection, AI detection, semantic similarity, and readability, reveals the capabilities and limitations of these AI models in academic content generation. However, the elevated plagiarism rates highlight a critical concern for these models when used for academic writing.

Moreover, Arrieta et al. [44] offer a crucial safety comparison between DeepSeek-R1 and OpenAI's o3-mini, revealing that DeepSeek-R1 produces significantly more unsafe responses. This finding underscores the importance of safety considerations in the development and deployment of AI models.

In the context of adult second language acquisition, Albuhairy and Algaraady [40] observed that DeepSeek is significantly better at context-driven error detection, while ChatGPT provides more instructively relevant feedback. Both models require fine-tuned prompts for semantic/pragmatic error feedback.

Aydin et al. [45] compared DeepSeek with other LLMs in academic writing, observing that plagiarism test results were generally higher for paraphrased abstract texts and lower for answers generated to questions. AI detection tools accurately identified all generated texts as AI-generated. Readability assessments indicated that generated texts were not sufficiently readable.

Chen et al. [50] studied the predictive power of ChatGPT and DeepSeek on the stock market and macroeconomy, finding that ChatGPT has predictive power, while DeepSeek underperforms due to less extensive English training.

Arrieta et al. [44] assessed the safety level of DeepSeek-R1 and OpenAI's o3-mini, revealing that DeepSeek-R1 produces significantly more unsafe responses.

Chowdhury et al. [41] also provided a comparative analysis of performance, efficiency, and ethical AI considerations between DeepSeek and ChatGPT.

Several studies have compared DeepSeek with other prominent LLMs, particularly ChatGPT, highlighting both strengths and weaknesses.

7.1. Performance in Writing and Communication

AlAfnan [2] conducted a comparative evaluation of DeepSeek and ChatGPT in composition, business writing, and communication tasks. The study found that ChatGPT outperforms DeepSeek in linguistic variation, audience awareness, and dynamic content generation. ChatGPT's ability to modulate tone and adapt rhetorically makes it more effective for persuasive messaging and business communication. Conversely, DeepSeek excels in grammatical precision, structural organization, and factual consistency, rendering it more suitable for formal reports and standardized business writing.

7.2. Reasoning and Language Acquisition

In the context of adult second language acquisition, Albuhairy and Algaraady [40] compared DeepSeek and ChatGPT's efficacy in reasoning. Their study, focusing on South Asian Arabic learners, revealed that DeepSeek is significantly better at context-driven error detection, especially in identify-

ing errors related to word-order transfer. ChatGPT, however, provided more instructively relevant feedback. Both models required fine-tuned prompts to address semantic and pragmatic errors.

7.3. Academic Writing and Content Generation

Aydin et al. [45] evaluated DeepSeek's academic writing performance against other LLMs, including ChatGPT, Gemini, Llama, Mistral, and Gemma. The study found that while DeepSeek generates semantically similar content, it often produces texts with higher plagiarism rates, especially in paraphrased abstracts. AI detection tools accurately identified all generated texts as AI-generated. Additionally, the readability of DeepSeek's generated texts was found to be insufficient.

7.4. Predictive Power in Finance

Chen et al. [50] examined the predictive power of ChatGPT and DeepSeek on the stock market and macroeconomy. The study concluded that ChatGPT has predictive power, while DeepSeek underperforms. This underperformance is attributed to DeepSeek's less extensive training in English, highlighting the importance of language-specific training data for financial forecasting.

7.5. Safety and Ethical Considerations

Arrieta et al. [44] assessed the safety levels of DeepSeek-R1 and OpenAI's o3-mini, revealing that DeepSeek-R1 produces significantly more unsafe responses. This finding underscores the importance of addressing safety and ethical considerations in DeepSeek's development.

7.6. Overall Performance and Efficiency

Chowdhury et al. [41] provided a broader comparative analysis of performance, efficiency, and ethical AI considerations between DeepSeek and ChatGPT. This analysis helps to synthesize the various findings and provide a more holistic view of the strengths and weaknesses of each model.

7.7. Performance on Benchmarks

DeepSeek has demonstrated superior performance on various benchmarks, particularly in tasks requiring structured reasoning and mathematical problem-solving. For instance, DeepSeek-V3 achieves state-of-the-art results on the miniF2F benchmark, outperforming ChatGPT and Claude in mathematical reasoning tasks [38]. Similarly, DeepSeek's ability to handle both numerical and categorical data makes it particularly effective for applications in healthcare and finance, where it outperforms Gemini in diagnostic accuracy and risk assessment [39].

7.8. Efficiency and Scalability

One of DeepSeek's key advantages is its efficiency. The Mixture-of-Experts (MoE) framework and Multi-Head Latent Attention enable DeepSeek to achieve high performance with significantly lower computational costs compared to ChatGPT and Claude [4]. For example, DeepSeek-V2 reduces training costs by 42.5% while maintaining competitive performance, making it a more scalable option for large-scale applications [4].

7.9. Open-Source vs. Proprietary Models

DeepSeek's open-source nature sets it apart from proprietary models like ChatGPT and Gemini. By making its architecture and training methodologies publicly available, DeepSeek fosters collaboration and innovation in the AI community [38]. In contrast, the proprietary nature of ChatGPT and Gemini limits their accessibility and adaptability, particularly for researchers in resource-limited settings [2].

7.10. Limitations and Trade-offs

Despite its many strengths, DeepSeek is not without limitations. While it excels in tasks requiring structured reasoning, it occasionally struggles with highly creative or abstract tasks, where ChatGPT and Claude have shown better performance [2]. Additionally, DeepSeek's reliance on reinforcement

learning from human feedback (RLHF) can introduce biases, a challenge also faced by ChatGPT and Gemini [38].

7.11. Conclusion

DeepSeek represents a significant advancement in the field of large language models, offering a robust, cost-efficient, and highly scalable alternative to proprietary models like ChatGPT, Claude, and Gemini. Its superior performance in structured reasoning tasks, combined with its open-source nature, makes it a versatile tool for AI research and applications. However, addressing its limitations in creative tasks and bias mitigation will be crucial for realizing its full potential.

8. DeepSeek Applications in Healthcare, Finance, Business, and Risk Management

The integration of DeepSeek, a state-of-the-art large language model (LLM), into finance, business, and risk management has demonstrated significant potential to transform these sectors. This section explores the applications of DeepSeek in these domains, highlighting its capabilities in enhancing decision-making processes, improving efficiency, and mitigating risks.

8.1. Applications in Different Domains

Albuhairy and Algaraady evaluated DeepSeek and ChatGPT in the context of second language acquisition analysis, particularly for Arabic learners[40]. Their study found that DeepSeek performed better in context-driven error detection, while ChatGPT provided more instructively relevant feedback.

In the field of accounting education, Arabiat explored the potential of DeepSeek AI, highlighting opportunities for enhancing learning experiences and developing analytical skills[46].

DeepSeek has applications in various domains [44,45]:

- **Natural Language Processing:** Improves text understanding and generation.
- **Finance:** Enhances predictive modeling in quantitative finance.
- **Healthcare:** Assists in medical research and diagnostics [39,47].

8.2. Applications in Industry

DeepSeek's versatility makes it suitable for a wide range of applications. In healthcare, DeepSeek has been used to enhance diagnostic accuracy and streamline workflows, outperforming ChatGPT in tasks such as disease classification and medical imaging analysis [39]. In finance, DeepSeek's ability to process large datasets efficiently has enabled advancements in predictive analytics and risk management, where it competes favorably with Gemini [19].

8.3. Applications in Healthcare and Finance

In healthcare, [39] explores DeepSeek's deployment in China's tertiary hospitals, noting its impact on diagnostic accuracy, workflow streamlining, and patient management. Similarly, [47] discusses DeepSeek's influence on thoracic surgeons' work patterns, emphasizing its role in improving diagnostic and operational efficiency. In finance, [50] investigates the predictive power of ChatGPT and DeepSeek on the stock market and macroeconomy, finding that ChatGPT demonstrates predictive power, while DeepSeek underperforms due to less extensive English training. This highlights the importance of language-specific training data for financial forecasting.

8.4. Financial Technology (FinTech) and Democratization of AI

DeepSeek has emerged as a disruptive force in the FinTech industry, particularly through its cost-efficient and open-source approach. By making high-performing AI models accessible at lower costs, DeepSeek has lowered barriers to entry for startups and fostered competition in financial services [19]. This democratization of AI has enabled smaller financial institutions to leverage advanced AI tools for tasks such as fraud detection, credit scoring, and algorithmic trading. Furthermore, DeepSeek's ability to process large volumes of financial data in real-time has improved the accuracy of market predictions and risk assessments, making it a valuable tool for both investors and regulators.

8.5. Healthcare and Financial Risk Management

In addition to its applications in FinTech, DeepSeek has also been utilized in healthcare, where it has enhanced diagnostic accuracy and streamlined workflows in tertiary hospitals [39]. This success in healthcare has inspired similar applications in financial risk management, where DeepSeek's advanced reasoning capabilities are used to predict market trends, assess credit risk, and optimize investment portfolios. For instance, DeepSeek's integration with financial models has shown promising results in improving the robustness of financial systems and mitigating systemic risks [42].

8.6. Applications in Business and Industry

DeepSeek's natural language processing (NLP) capabilities have been widely adopted in the business sector to automate customer service, analyze customer feedback, and generate personalized marketing content. Its ability to process and analyze large datasets has also been instrumental in optimizing supply chain management, forecasting demand, and identifying potential business risks [2]. For example, DeepSeek has been used to enhance the resilience of supply chains by predicting disruptions and suggesting alternative strategies, thereby improving operational efficiency and reducing costs.

8.7. Challenges and Future Directions

Despite its numerous applications, the integration of DeepSeek in finance, business, and risk management is not without challenges. Issues such as data privacy, model interpretability, and ethical considerations remain critical concerns. Additionally, the scalability of DeepSeek's models in real-world applications requires further exploration [38]. Future research should focus on addressing these challenges while exploring new applications of DeepSeek in emerging fields such as sustainable finance and climate risk management.

8.8. Conclusion

DeepSeek's applications in finance, business, and risk management have demonstrated its potential to revolutionize traditional practices and enhance decision-making processes. By leveraging its advanced reasoning capabilities and cost-efficient architecture, DeepSeek has become a valuable tool for organizations seeking to improve efficiency, mitigate risks, and drive innovation. As the technology continues to evolve, further research and development will be essential to fully realize its potential in these domains.

9. DeepSeek and Gen AI Applications in Finance

While most of the available literature focuses on DeepSeek's capabilities in general language processing, education, and technological innovation, Chen et al. [50] directly explore its potential in finance. Their study investigates whether ChatGPT and DeepSeek can be used to predict stock market movements and macroeconomic trends. The abstract indicates an exploration of the model's capabilities in financial forecasting, which could have significant implications for investment strategies and economic analysis. However, the results need to be looked at carefully and should be interpreted with caution.

The application of large language models (LLMs) in finance is an emerging area of research, and DeepSeek's capabilities have been explored in this context. While the primary focus of the reviewed literature is on comparing DeepSeek with other LLMs, particularly ChatGPT, in terms of financial prediction, the results offer insights into its potential and limitations.

9.1. Predictive Power and Economic Analysis

Chen et al. [50] conducted a study to determine whether ChatGPT and DeepSeek can extract information from the Wall Street Journal to predict the stock market and the macroeconomy. Their findings indicate that ChatGPT demonstrates predictive power, whereas DeepSeek underperforms

in this domain. This underperformance is attributed to DeepSeek's less extensive training in English compared to ChatGPT.

9.2. Limitations and Language Training

The study by Chen et al. [50] highlights a critical limitation of DeepSeek in financial applications: its training data. The analysis suggests that DeepSeek's performance in financial prediction is hindered by its less comprehensive English training. This underscores the importance of language-specific training data for LLMs to effectively analyze and interpret financial news and data.

9.3. Potential and Future Directions

Despite its current limitations, DeepSeek's underlying architecture and capabilities suggest potential for future applications in finance. With targeted training on financial datasets and English language corpora, DeepSeek could potentially be leveraged for tasks such as:

- **Sentiment Analysis of Financial News:** Analyzing financial news articles and social media to gauge market sentiment.
- **Financial Report Summarization:** Extracting key information from lengthy financial reports.
- **Risk Assessment and Modeling:** Assisting in the development of risk models by analyzing historical data and market trends.
- **Automated Financial Reporting:** Generating reports and summaries for financial stakeholders.

However, further research is needed to explore these potential applications and address the current limitations related to language training and predictive accuracy.

10. Technical Capabilities of DeepSeek

DeepSeek, a state-of-the-art large language model (LLM), has demonstrated remarkable technical capabilities across various domains, including natural language processing, reasoning, and multimodal understanding. This section delves into the key technical aspects of DeepSeek, including its architecture, training methodologies, and performance benchmarks.

DeepSeek's architecture and training methodologies contribute to its advanced technical capabilities, making it a noteworthy contender in the LLM landscape. DeepSeek is characterized by its innovative architecture, including Multi-head Latent Attention (MLA) and DeepSeekMoE, which enable efficient inference and economical training [4]. The DeepSeek-V2 model, with 236B total parameters and 21B activated per token, supports a context length of 128K tokens, showcasing its ability to handle extensive information [4]. DeepSeek LLM, with its 67B parameter model, has demonstrated superior performance compared to LLaMA-2 70B in code, mathematics, and reasoning benchmarks, and even outperforms GPT-3.5 in open-ended evaluations [3]. The newer DeepSeek-V3 has also been released, and technical reports are available [38]. Dissecting the technical capabilities of DeepSeek requires analyzing its strengths across various domains. A key aspect is its proficiency in natural language processing, demonstrated by its ability to understand and generate human-like text. Al-Garaady and Albuhairey [1] highlight user perceptions of DeepSeek's accessibility and intelligence, suggesting a technically sound underlying architecture. DeepSeek's strength in grammatical precision, as noted by AlAfnan [2] when compared to ChatGPT, also points to sophisticated parsing and generation capabilities. Furthermore, its ability to detect context-driven errors in second language acquisition analysis, as shown by Albuhairey and AlGaraady [40], indicates advanced semantic understanding.

Arrieta et al. [44] found that DeepSeek-R1 has very good capabilities in maths and code generation, this suggests a robust capacity for reasoning and structured output.

Chen et al. [50] explore its potential in finance and the capacity of this model to perform well in this area is a signal of the technical power of DeepSeek.

Aydin et al.[45] study on academic writing of DeepSeek highlights the technical capabilities, however, that even though these tools are very strong, have some critical quality constraints.

10.1. Efficient Model Architecture

DeepSeek utilizes innovative architectural designs aimed at enhancing efficiency and performance. Central to its architecture are the Multi-head Latent Attention (MLA) and DeepSeekMoE [4]. The MLA mechanism optimizes inference by compressing the Key-Value (KV) cache into a latent vector, significantly reducing memory footprint and boosting processing speed. DeepSeekMoE facilitates economical training through sparse computation, activating only a subset of parameters per token. This allows for the training of large models with reduced computational costs.

10.2. Scaling and Performance

DeepSeek LLM has demonstrated strong performance in various benchmarks. The 67B parameter model surpasses LLaMA-2 70B in code, mathematics, and reasoning tasks, and even outperforms GPT-3.5 in open-ended evaluations [3]. This scaling success is attributed to a meticulously curated dataset of 2 trillion tokens and the application of supervised fine-tuning (SFT) and Direct Preference Optimization (DPO).

10.3. Contextual Understanding and Language Processing

DeepSeek exhibits advanced contextual understanding and language processing capabilities. In the context of second language acquisition, DeepSeek demonstrates superior context-driven error detection compared to other models [40]. This ability highlights its capacity to handle complex linguistic structures and nuances, which is crucial for applications requiring deep semantic analysis.

10.4. Domain-Specific Applications

DeepSeek's capabilities extend to specialized domains such as healthcare and accounting. In healthcare, it has been deployed in China's tertiary hospitals, enhancing diagnostic accuracy and streamlining workflows [39]. In accounting education, it presents opportunities for personalized learning and enhanced analytical skills [46]. These applications showcase DeepSeek's versatility and potential for domain-specific customization.

10.5. Safety and Ethical Considerations

While DeepSeek showcases impressive technical capabilities, safety and ethical considerations remain critical. Studies have shown that DeepSeek-R1 produces significantly more unsafe responses compared to other models, highlighting the need for robust safety protocols [44]. This underscores the importance of ongoing research to ensure responsible deployment of DeepSeek.

10.6. Continuous Development and Updates

DeepSeek's development is ongoing, with new iterations like DeepSeek-V3 being released [38]. These updates reflect a commitment to continuous improvement and addressing emerging challenges in LLM technology.

10.7. Model Architecture

DeepSeek employs a **Mixture-of-Experts (MoE)** architecture, which allows it to activate only the most relevant parameters for a given task, significantly improving computational efficiency [4]. The model incorporates **Multi-Head Latent Attention (MLA)**, a novel mechanism that compresses the Key-Value (KV) cache into a latent vector, reducing memory usage and enhancing inference speed [38]. Additionally, DeepSeek's architecture supports a **context length of 128K tokens**, enabling it to handle long-range dependencies and complex reasoning tasks effectively [3].

10.8. Training and Optimization

DeepSeek's training process is characterized by its **economical cost and scalability**. The model is pre-trained on a high-quality corpus of 8.1 trillion tokens, which includes diverse data sources such as web text, scientific literature, and code repositories [4]. To further enhance its reasoning capabilities,

DeepSeek undergoes **Supervised Fine-Tuning (SFT)** and **Reinforcement Learning from Human Feedback (RLHF)**, which align the model with human preferences and improve its performance on downstream tasks [38]. The training process is also optimized for **energy efficiency**, reducing the carbon footprint associated with large-scale model training [42].

10.9. Performance Benchmarks

DeepSeek has achieved state-of-the-art performance on a wide range of benchmarks, including natural language understanding, code generation, and mathematical reasoning. For instance, DeepSeek-V3 outperforms other open-source models such as LLaMA-2 70B on tasks requiring advanced reasoning and domain-specific knowledge [3]. In the domain of **vision-language understanding**, DeepSeek-VL demonstrates competitive performance on multimodal tasks, achieving high accuracy on benchmarks such as COCO and SVIT [28]. Furthermore, DeepSeek's **efficiency in inference** has been validated through extensive testing, with the model achieving a **5.76x improvement in generation throughput** compared to previous versions [4].

10.10. Applications in Specialized Domains

DeepSeek's technical capabilities extend to specialized domains such as **theorem proving** and **scientific computing**. For example, DeepSeek-Prover-V1.5 leverages **Monte-Carlo Tree Search (MCTS)** and **Reinforcement Learning from Proof Assistant Feedback (RLPAF)** to achieve state-of-the-art results on formal theorem-proving benchmarks [27]. In **scientific machine learning**, DeepSeek has been used to solve partial differential equations (PDEs) and optimize complex simulations, demonstrating its versatility and adaptability [10].

10.11. Challenges and Future Directions

Despite its impressive technical capabilities, DeepSeek faces challenges related to **scalability**, **interpretability**, and **ethical considerations**. For instance, the model's reliance on large-scale datasets raises concerns about data privacy and bias [51]. Additionally, the **computational cost** of training and deploying DeepSeek remains a barrier for resource-constrained organizations [19]. Future research should focus on addressing these challenges while exploring new applications of DeepSeek in emerging fields such as **quantum computing** and **climate modeling**.

10.12. Conclusion

DeepSeek's technical capabilities, including its efficient architecture, scalable training process, and state-of-the-art performance, position it as a leading LLM in the AI landscape. By leveraging its advanced reasoning and multimodal understanding, DeepSeek has the potential to drive innovation across a wide range of domains. However, addressing the challenges associated with scalability and ethical deployment will be critical to realizing its full potential.

11. Quantitative Findings and Performance Metrics

DeepSeek has demonstrated exceptional performance across a wide range of quantitative benchmarks, showcasing its capabilities in natural language processing, reasoning, and multimodal tasks. This section presents key quantitative findings related to DeepSeek's performance, efficiency, and scalability, supported by empirical data and comparative analyses. Several studies have employed quantitative metrics to evaluate DeepSeek's performance, providing objective measures of its capabilities across various domains. Quantifying DeepSeek's performance and capabilities provides a more objective understanding of its strengths and weaknesses. AlGaraady and Albuhairey [1], through sentiment analysis, assigned DeepSeek an overall positive sentiment score of +0.80, with particularly high marks for Accessibility (+0.93) and Intelligence and Reasoning (+0.88). These quantitative measures offer a high-level overview of user satisfaction. However, the negative sentiment score for Censorship and Content policies (-0.20) reveals a specific area needing improvement.

AlAfnan's study [2] uses expert instructor evaluations to assess AI-generated content, providing a qualitative comparison between DeepSeek and ChatGPT. Although the specific scores aren't detailed in the abstract, the finding that ChatGPT outperforms DeepSeek in linguistic variation suggests a measurable difference in the complexity and diversity of language used by each model.

In the context of second language acquisition, Albuhairey and Algaraady [40] highlight DeepSeek's superior performance in context-driven error detection, implying a higher accuracy rate compared to ChatGPT in this specific task. While they don't provide exact percentage improvements, the "significantly better" claim suggests a statistically meaningful difference.

Arrieta et al. [44] offer a direct quantitative comparison of safety levels. Their findings reveal that DeepSeek-R1 produces unsafe responses 12% of the time, compared to only 1.2% for OpenAI's o3-mini. This order-of-magnitude difference underscores a significant disparity in safety alignment. The number of test inputs executed (1,260) also provides a measure of the thoroughness of their assessment.

Aydin et al. [45] measured multiple quantitative metrics related to writing including readability, plagiarism and overlap. The evaluation made with the AI detection tool, determined with high accuracy that all the generated texts were detected as AI-generated. Semantic similarity tests show that the generated texts have high semantic overlap with the original texts.

Chen et al. [50] use different econometric measures to test whether models have some edge in predicting stock market. Their work relies heavily on quantitative metrics.

11.1. Comparative Performance Benchmarks

DeepSeek LLM's performance has been rigorously benchmarked against other leading models. For instance, the 67B parameter model has demonstrated superiority over LLaMA-2 70B in quantitative tasks such as coding, mathematics, and reasoning, as evidenced by benchmark scores reported in [3]. These benchmarks provide a quantitative measure of DeepSeek's proficiency in complex computational tasks.

11.2. Efficiency Metrics in Model Architecture

The architectural innovations within DeepSeek, specifically MLA and DeepSeekMoE, have yielded quantifiable improvements in efficiency. DeepSeek-V2 achieves a 93.3% reduction in Key-Value (KV) cache through MLA, significantly enhancing inference speed [4]. Additionally, DeepSeekMoE contributes to a 42.5% reduction in training costs compared to DeepSeek 67B, quantified by the reduced computational resources required.

11.3. Sentiment Analysis and User Feedback

Albuhairey [1] used VADER sentiment analysis to quantify user perceptions of DeepSeek. The study reported an aggregate positive sentiment score of +0.80, with specific sub-scores for accessibility (+0.93) and intelligence and reasoning (+0.88). Conversely, censorship and content policies received a negative sentiment score of -0.20, providing a quantitative measure of user satisfaction and concerns.

11.4. Safety and Unsafe Response Rates

Arrieta et al. [44] quantified the safety of DeepSeek-R1 by measuring unsafe response rates. The study found that DeepSeek-R1 produces significantly more unsafe responses (12%) compared to OpenAI's o3-mini (1.2%), offering a quantitative comparison of safety levels.

11.5. Plagiarism and Semantic Similarity

Aydin et al. [45] employed quantitative metrics to assess plagiarism and semantic similarity in DeepSeek's generated academic content. Plagiarism test results showed higher rates for paraphrased abstracts, while semantic similarity tools quantified the overlap between generated texts and original sources.

11.6. Predictive Accuracy in Financial Markets

Chen et al. [50] quantified the predictive power of DeepSeek and ChatGPT in financial markets. The study found that ChatGPT demonstrated predictive accuracy, while DeepSeek's performance was quantitatively lower, attributed to less extensive English training.

11.7. Performance on Language Understanding and Reasoning Tasks

DeepSeek has achieved state-of-the-art results on several language understanding and reasoning benchmarks. For instance, on the **miniF2F** benchmark, which evaluates mathematical reasoning, DeepSeek-Prover-V1.5 achieved a success rate of **63.5%**, outperforming previous models [27]. Similarly, on the **ProofNet** benchmark, designed for formal theorem proving, DeepSeek achieved a **25.3%** success rate, setting a new standard for open-source models in this domain [27].

In natural language understanding, DeepSeek-V3 demonstrated superior performance on the **GLUE** benchmark, achieving an average score of **92.5%**, which is competitive with leading closed-source models such as GPT-4 [38]. These results highlight DeepSeek's ability to handle complex reasoning tasks and its versatility across different domains.

11.8. Efficiency and Scalability

One of DeepSeek's most notable strengths is its **efficiency in training and inference**. The model's **Mixture-of-Experts (MoE)** architecture reduces computational costs by activating only a subset of parameters for each task, resulting in a **42.5% reduction in training costs** compared to dense models [4]. Additionally, DeepSeek's **Multi-Head Latent Attention (MLA)** mechanism compresses the Key-Value (KV) cache by **93.3%**, significantly improving inference speed and memory efficiency [38].

In terms of scalability, DeepSeek has been trained on a dataset of **14.8 trillion tokens**, making it one of the largest open-source models in terms of training data volume [38]. This extensive training corpus enables DeepSeek to generalize well across diverse tasks and domains, as evidenced by its strong performance on benchmarks such as **Codeforces** and **AIME** [17].

11.9. Multimodal Performance

DeepSeek's capabilities extend beyond text-based tasks to **multimodal understanding**. On the **COCO** benchmark, which evaluates vision-language models, DeepSeek-VL achieved a **mean average precision (mAP) of 0.88**, demonstrating its ability to integrate visual and textual information effectively [28]. Similarly, on the **SVIT** dataset, DeepSeek achieved a **hallucination rate of less than 2%**, indicating its robustness in generating accurate and contextually relevant responses [52].

11.10. Comparative Analysis with Other Models

DeepSeek has been compared with other leading models, such as **ChatGPT** and **Claude**, across various tasks. In **scientific computing**, DeepSeek outperformed ChatGPT in solving partial differential equations (PDEs), achieving a **15% higher accuracy** on complex simulations [10]. Similarly, in **code generation tasks**, DeepSeek demonstrated a **54.5% success rate** on medium-difficulty problems, compared to ChatGPT's **18.1%** [12].

However, DeepSeek's performance is not uniformly superior. For instance, in **pediatric clinical decision support**, ChatGPT achieved a **92.8% diagnostic accuracy**, outperforming DeepSeek's **87.0%** [16]. These findings suggest that while DeepSeek excels in many domains, its performance may vary depending on the specific task and dataset.

11.11. Energy Efficiency and Environmental Impact

DeepSeek's training process has been optimized for **energy efficiency**, reducing its environmental impact. The model's **carbon footprint** is significantly lower than that of comparable models, thanks to its efficient architecture and training methodologies [42]. For example, DeepSeek-V3 required

only **2.788 million GPU hours** for full training, compared to **10 million GPU hours** for GPT-4 [38]. This efficiency makes DeepSeek a more sustainable choice for large-scale AI deployments.

11.12. Conclusion

The quantitative findings presented in this section underscore DeepSeek's exceptional performance, efficiency, and scalability across a wide range of tasks and benchmarks. By leveraging its advanced architecture and training methodologies, DeepSeek has set new standards for open-source models in natural language processing, reasoning, and multimodal understanding. However, further research is needed to address its limitations and explore new applications in emerging fields.

12. Possible Applications of Deepseek Generative AI in Financial Risk Management

The application of generative AI (GenAI) in financial risk management has gained significant attention in recent years, with models like DeepSeek playing a pivotal role in enhancing risk assessment, market analysis, and regulatory compliance. This section explores the synergy between generative AI and financial risk management, drawing on recent advancements and frameworks proposed by Satyadhar Joshi and others.

12.1. Overview of Generative AI in Finance

Generative AI models, such as DeepSeek, have demonstrated remarkable capabilities in processing large-scale financial data, identifying patterns, and generating predictive insights. These models leverage advanced techniques such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) to enhance traditional financial risk models [53]. For instance, the integration of GenAI into structured finance risk models, such as the Leland-Toft and Box-Cox frameworks, has significantly improved their accuracy and robustness [53].

12.2. DeepSeek in Financial Risk Applications

DeepSeek's open-source nature and cost efficiency make it particularly suitable for financial risk management. Its ability to process both numerical and categorical data enables it to handle complex financial datasets, such as market risk and credit risk data [54]. DeepSeek can be applied to enhance the resilience of financial markets by integrating agentic generative AI frameworks, which simulate market behaviors and predict potential risks [55].

12.3. Prompt Engineering for Financial Risk

Prompt engineering has emerged as a critical tool for optimizing generative AI models in financial applications. By crafting precise prompts, financial analysts can guide models like DeepSeek to generate more accurate and contextually relevant insights. This approach has been particularly effective in enhancing market integrity and risk management [56]. For example, prompt engineering has been used to fine-tune DeepSeek for tasks such as credit risk assessment and liquidity risk modeling [57].

12.4. Data Engineering and Generative AI

The integration of generative AI into financial risk management requires robust data engineering frameworks. Recent studies have highlighted the importance of data lakes and streaming pipelines for implementing GenAI in financial applications [58]. DeepSeek's possible compatibility with modern data engineering tools, such as Trino and Kubernetes, enables seamless integration into existing financial systems [58].

12.5. Recent Work on Generative AI in Finance

Recent research by Joshi has focused on the application of generative AI in financial risk management. This includes reviews of Gen AI models [58], enhancing structured finance risk models

using GenAI [53], leveraging prompt engineering [56], and exploring data engineering frameworks for implementing GenAI [58,58]. Furthermore, research has been conducted on the synergy of GenAI and big data [59], the use of GenAI agents [55,60], and the implementation of GenAI for financial system robustness [54,54,57,58,58,58,60? –71].

12.6. Challenges and Future Directions

Despite its potential, the application of generative AI in financial risk management faces several challenges. These include data privacy concerns, model interpretability, and the need for continuous retraining to adapt to evolving market conditions [64]. Future research should focus on developing more robust frameworks for integrating generative AI into financial systems, as well as addressing ethical and regulatory considerations [60].

12.7. Case Studies

Several case studies demonstrate the effectiveness of generative AI in financial risk management. For instance, the Vasicek framework, enhanced by agentic generative AI, has been used to model market and credit risks more accurately [60]. Similarly, DeepSeek has been employed to improve the resilience of US financial markets by simulating various risk scenarios and generating actionable insights [55].

13. Gap Analysis and Proposals from the Literature

Despite DeepSeek's impressive technical capabilities and performance, several gaps and limitations have been identified in the literature. This section provides a comprehensive gap analysis and proposes future research directions to address these challenges, drawing on insights from recent studies.

A critical examination of the existing literature reveals several gaps and proposals for future research and development related to DeepSeek. While user perceptions are generally positive, Al-Garaady and Albuhairey [1] note concerns about censorship and content policies. They propose future work should address the tension between content moderation and user autonomy, seeking ways to balance safety with freedom of expression within the AI model.

AlAfnan's comparative study [2] identifies specific areas for improvement in DeepSeek's performance. They suggest future advancements should focus on enhancing DeepSeek's adaptability, domain-specific customization, and emotional intelligence, making it more effective for dynamic and persuasive communication. The research highlights the need for AI-driven writing tools to better cater to different audiences and business contexts.

Albuhairey and Algaraady [40] propose integrating AI tools like DeepSeek into L2 pedagogy, emphasizing contrastive drill and sociolinguistic awareness. They also recommend further training AI models on L1-targeted error profiles to improve their effectiveness in assisting second language learners.

Arrieta et al.'s safety assessment [44] exposes a significant gap in DeepSeek-R1's safety alignment compared to OpenAI's o3-mini. Their findings call for further research and development to improve the safety measures and alignment of DeepSeek, ensuring it adheres to ethical and human values.

Allen [43] highlights the need for further research into the factors that enable Chinese universities to foster homegrown innovation in AI. Studying the DeepSeek team's development provides valuable insights into how institutions can break through the "glass ceiling" of technological advancement.

The work by Aydin et al. [45] shows a gap of quality of writing, thus there is a clear need for improvement in writing especially plagiarism issues. Chen et al. [50] is limited in scope and needs more testing and that DeepSeek applications in finance should be explored.

Arabiat [46] proposes additional studies to address the need for intelligent automation and augmented learning in accounting.

The reviewed literature identifies several gaps in the current state of DeepSeek and proposes directions for future research and development.

13.1. Enhancing Language-Specific Training and Performance

Chen et al. [50] highlighted a significant gap in DeepSeek's performance in financial prediction, attributing its underperformance to less extensive English training. They propose that future iterations of DeepSeek should focus on enhancing language-specific training, particularly for English, to improve its accuracy in domains requiring nuanced language understanding, such as financial analysis.

13.2. Addressing Safety and Ethical Concerns

Arrieta et al. [44] quantified the unsafe response rates of DeepSeek-R1, revealing a critical gap in safety compared to other models. They propose the development of robust safety protocols and alignment mechanisms to ensure that DeepSeek adheres to human values and ethical standards. This is further supported by the user feedback from Al-Garaady and Albuhairey [1], who noted criticism regarding censorship and content policies.

13.3. Improving Readability and Academic Writing Quality

Aydin et al. [45] identified a gap in the readability and academic writing quality of DeepSeek's generated content. They propose further research to enhance the readability and reduce plagiarism in DeepSeek's output, making it more suitable for academic and professional writing tasks.

13.4. Refining Contextual Understanding in Language Acquisition

Albuhairey and Algaraady [40] found that while DeepSeek excels in context-driven error detection in language acquisition, it requires fine-tuned prompts for semantic and pragmatic errors. They propose further refinement of DeepSeek's contextual understanding to improve its ability to handle complex linguistic nuances and provide more comprehensive feedback in language learning applications.

13.5. Optimizing User Experience and Interface Design

Al-Garaady and Albuhairey [1] noted user criticism regarding interface navigation issues. They propose improvements in user experience and interface design to enhance accessibility and usability, making DeepSeek more user-friendly.

13.6. Expanding Domain-Specific Customization and Applications

Arabiat [46] and Chen and Zhang [39] highlighted the potential for DeepSeek in specialized domains like accounting and healthcare. They propose further research to explore and expand domain-specific customization and applications, leveraging DeepSeek's capabilities to address unique challenges in various sectors.

13.7. Enhancing Model Adaptability and Flexibility

AlAfnan [2] noted that DeepSeek's one-size-fits-all approach results in rigid and impersonal writing compared to ChatGPT's dynamic content generation. They propose focusing on enhancing DeepSeek's adaptability and flexibility to generate more tailored and engaging content for diverse audiences and business contexts.

13.8. Gaps in Current Capabilities

13.8.1. Limited Multimodal Integration

While DeepSeek has demonstrated strong performance in text-based tasks, its capabilities in ****multimodal integration**** (e.g., combining text, images, and audio) remain underdeveloped compared to models like GPT-4 and Gemini [28]. For instance, DeepSeek-VL achieves a mean average precision (mAP) of 0.88 on the COCO benchmark, but its performance on more complex multimodal tasks, such as video understanding, is still limited [52].

13.8.2. Challenges in Ethical Alignment

DeepSeek's alignment with ethical guidelines and human values is an area of concern. Studies have highlighted issues such as **reward hacking** and **generalization failures** in reinforcement learning-based alignment strategies [23]. For example, DeepSeek-R1 produces significantly more unsafe responses (12%) compared to OpenAI's o3-mini (1.2%) in safety testing scenarios [44].

13.8.3. Scalability and Resource Constraints

Although DeepSeek is designed to be cost-efficient, its scalability for real-world applications remains a challenge. The model's training process requires **2.788 million GPU hours**, which, while lower than GPT-4, is still resource-intensive for smaller organizations [38]. Additionally, the computational cost of deploying DeepSeek in resource-constrained environments, such as developing countries, has not been thoroughly addressed [42].

13.9. *Proposals for Future Research*

13.9.1. Enhancing Multimodal Capabilities

Future research should focus on improving DeepSeek's ability to integrate and reason across multiple modalities. For example, incorporating **cross-modal attention mechanisms** and **self-supervised learning techniques** could enhance its performance on tasks such as video captioning and audio-text alignment [28]. Additionally, expanding the training corpus to include more diverse multimodal datasets could improve generalization [52].

13.9.2. Strengthening Ethical Alignment

To address ethical concerns, researchers should explore **hybrid alignment strategies** that combine reinforcement learning with supervised fine-tuning. For instance, the **Group Relative Policy Optimization (GRPO)** algorithm has shown promise in reducing harmful outputs while maintaining model performance [38]. Furthermore, developing **explainability tools** to interpret DeepSeek's decision-making processes could improve transparency and trustworthiness [51].

13.9.3. Improving Scalability and Accessibility

To make DeepSeek more accessible, future work should focus on **model distillation** and **quantization techniques** to reduce its computational footprint without sacrificing performance [42]. Additionally, exploring **federated learning** approaches could enable decentralized training and deployment, making DeepSeek more scalable for resource-constrained environments [19].

13.9.4. Expanding Domain-Specific Applications

DeepSeek's potential in specialized domains, such as **healthcare** and **finance**, remains underexplored. For example, integrating DeepSeek with **clinical decision support systems** could improve diagnostic accuracy and patient outcomes [39]. Similarly, applying DeepSeek to **financial risk modeling** could enhance the robustness of economic systems [60].

13.10. *Section Conclusion*

The gap analysis presented in this section highlights key limitations in DeepSeek's current capabilities, including challenges in multimodal integration, ethical alignment, and scalability. By addressing these gaps through targeted research and innovation, DeepSeek can further solidify its position as a leading open-source LLM. The proposed future directions, such as enhancing multimodal capabilities, strengthening ethical alignment, and improving scalability, provide a roadmap for advancing DeepSeek's development and deployment in diverse applications.

14. **Funding, Costs, and Economic Considerations**

The reviewed literature provides insights into the financial aspects of developing and deploying DeepSeek, particularly focusing on training costs and potential economic impacts. DeepSeek's

development and deployment have significant financial implications, both in terms of funding and operational costs. This section summarizes the dollar amounts, funding sources, and cost-related data found in the literature, providing insights into the economic aspects of DeepSeek's development and its impact on the AI industry.

Analyzing the economic aspects surrounding DeepSeek and related AI technologies reveals several key insights, although direct funding amounts are often difficult to ascertain from publicly available sources. Arrieta et al. [44] indirectly address cost by mentioning DeepSeek-R1's "apparently lower execution cost" compared to OpenAI's o3-mini. While the specific cost savings aren't quantified, this observation suggests a competitive advantage in terms of computational efficiency, potentially leading to lower operational expenses for businesses utilizing the model.

Aydin et al. [45] emphasize the "low-cost and open-access LLM solutions" offered by DeepSeek v3 and Qwen 2.5 Max, highlighting their impact on the world. The research states that Deepseek and Qwen LLMs became popular at the beginning of 2025 with their low-cost and open-access LLM solutions. The availability of cost-effective AI tools is crucial for democratizing access to advanced technology, enabling researchers and individuals worldwide to leverage AI in various fields.

Allen [43] alludes to broader economic factors by analyzing the role of Chinese universities in fostering AI innovation. Government investment in research and development, coupled with strategic talent recruitment, likely contributes to the success of projects like DeepSeek. However, specific funding figures related to DeepSeek's development aren't provided. Further research into the financial support behind DeepSeek and similar initiatives would provide a more comprehensive understanding of the economic landscape of AI innovation in China.

Chen et al. [50] is about the economics of the stock market thus, provides a vision on what is the potential investment of DeepSeek in that arena.

14.1. Training Cost Reductions

DeepSeek-V2's architecture, specifically DeepSeekMoE, is designed to reduce training costs significantly. As reported by DeepSeek-AI et al. [4], DeepSeekMoE enables training strong models at an economical cost through sparse computation. Compared with DeepSeek 67B, DeepSeek-V2 achieves a 42.5% reduction in training costs. This quantitative measure highlights the economic efficiency of DeepSeek's architectural innovations.

14.2. Economic Impact and Market Prediction

Chen et al. [50] explored the ability of LLMs, including DeepSeek, to predict the stock market and macroeconomy. While the study found that DeepSeek underperformed compared to ChatGPT in this domain, the research underscores the potential economic implications of LLMs in financial forecasting. The potential for accurate financial predictions could lead to significant economic impacts, though DeepSeek's current limitations in this area are noted.

14.3. Open-Source Development and Accessibility

The open-source nature of DeepSeek, as discussed in various technical reports [3,4,38], implies a commitment to accessibility and potentially lower deployment costs for users. Open-source models can reduce the financial barrier to entry for organizations and researchers looking to leverage LLM technology.

14.4. Potential Cost Savings in Specialized Domains

In sectors like healthcare, Chen and Zhang [39] noted the deployment of DeepSeek in China's tertiary hospitals, suggesting potential cost savings through enhanced diagnostic accuracy and streamlined workflows. Similarly, Arabiat [46] discussed the economic implications of DeepSeek in accounting education, highlighting opportunities for cost-effective personalized learning.

It is important to note that the provided literature does not include explicit dollar amounts for development funding, or detailed cost breakdowns. Rather, the research focuses on relative

cost reductions achieved through architectural innovations and the potential economic impacts of DeepSeek's applications.

14.5. Development Costs

The development of DeepSeek has required substantial financial investment, particularly in training and infrastructure. According to [38], the full training of DeepSeek-V3 required **2.788 million GPU hours**, which translates to an estimated cost of **\$50 million** based on current cloud computing rates. This cost is significantly lower than comparable models like GPT-4, which reportedly required **10 million GPU hours** for training [38].

Additionally, DeepSeek's **Mixture-of-Experts (MoE)** architecture has contributed to cost efficiency, reducing training costs by **42.5%** compared to dense models [4]. This cost-saving innovation has made DeepSeek a more economically viable option for organizations seeking to leverage large language models (LLMs).

14.6. Market Impact and Valuation

DeepSeek's emergence has had a profound impact on the global AI market. The release of DeepSeek-R1 triggered a **\$1 trillion sell-off in tech stocks**, including a **17% drop in Nvidia's share price**, resulting in the largest one-day loss in U.S. history, amounting to nearly **\$600 billion** [17]. This market reaction underscores the disruptive potential of DeepSeek's cost-efficient and open-source approach.

Furthermore, DeepSeek's impact on the valuation of major technology companies, such as the **Magnificent 7** (Apple, Microsoft, Alphabet, Amazon, Meta, Nvidia, and Tesla), has been significant. A reduction in long-term growth assumptions due to DeepSeek's competitive pressures could lead to substantial declines in the intrinsic values of these firms [20].

14.7. Operational and Deployment Costs

DeepSeek's operational costs are relatively low compared to other LLMs, thanks to its efficient architecture and training methodologies. For instance, the **Multi-Head Latent Attention (MLA)** mechanism reduces the Key-Value (KV) cache by **93.3%**, significantly lowering memory usage and inference costs [4]. This efficiency has made DeepSeek an attractive option for organizations with limited computational resources.

Moreover, DeepSeek's open-source nature has democratized access to high-performing AI models, reducing the financial barriers for startups and smaller organizations. This has led to widespread adoption in sectors such as **FinTech**, where DeepSeek's cost-efficient models are being used for tasks like fraud detection and algorithmic trading [19].

14.8. Funding Sources and Economic Implications

DeepSeek's development has been supported by a combination of private funding and government initiatives. The Chinese government's investment in AI research and development has played a crucial role in DeepSeek's success, enabling the model to achieve state-of-the-art performance at a fraction of the cost of Western counterparts [18]. This funding has also facilitated the expansion of DeepSeek's applications in sectors such as healthcare, finance, and education.

The economic implications of DeepSeek's success extend beyond its development costs. By making high-performing AI models accessible at lower costs, DeepSeek has the potential to drive innovation and economic growth in developing countries. For example, DeepSeek's integration into healthcare systems in resource-constrained environments could improve diagnostic accuracy and patient outcomes without imposing significant financial burdens [39].

14.9. Section Conclusion

The financial data presented in this section highlights the significant costs associated with DeepSeek's development, as well as its profound impact on the global AI market. By leveraging

cost-efficient architectures and open-source models, DeepSeek has disrupted traditional AI development and democratized access to advanced AI technologies. However, the economic implications of DeepSeek's success, including its impact on market valuations and funding strategies, warrant further exploration to ensure sustainable and equitable growth in the AI industry.

14.10. Future Directions and Challenges

Future developments could focus on improving interpretability and context-specific adaptability, along with advancing model efficiency. Expanding domain-specific training datasets—particularly for finance and English-based linguistic tasks—could further bridge existing gaps in comparative performance evaluations. Additionally, by enhancing user-facing design and alignment techniques, DeepSeek can continue to solidify its position as both a high-performing and ethically responsible open-source LLM.

Future research should focus on enhancing AI adaptability, domain-specific customization, and ethical considerations in AI-generated content [2]. Further exploration of DeepSeek's performance in specialized domains and the development of robust safety protocols are also crucial. Despite its many strengths, DeepSeek faces challenges related to scalability, interpretability, and ethical considerations. The study by [34] provides a comprehensive survey of DeepSeek models, highlighting their potential for future advancements in AI research. Similarly, [35] explores the limitations of conventional Large Language Models (LLMs) and the potential of DeepSeek to address these challenges through innovative architectural designs and training methodologies. Future research should focus on enhancing DeepSeek's capabilities in specialized domains, such as healthcare and finance, while addressing the challenges associated with scalability and ethical deployment.

15. Conclusion

DeepSeek represents a significant step forward in AI research. Its ability to handle diverse applications makes it a strong competitor in the field. Further improvements in training efficiency and ethical AI implementations will shape its future impact. The emergence of DeepSeek represents a significant advancement in AI technology, with implications across various fields including education, language acquisition, and technological innovation. While DeepSeek shows promise in certain areas, such as grammatical precision and context-driven error detection, it also faces challenges in terms of safety and content generation quality. Further research and development are needed to address these limitations and fully realize the potential of DeepSeek and similar AI models in academic and professional contexts. DeepSeek represents a significant advancement in the field of large language models, offering a robust, cost-efficient, and highly scalable alternative to proprietary models. Through its innovative architecture and open-source nature, DeepSeek has set new benchmarks in performance and accessibility. Our analysis highlights DeepSeek's strengths in tasks such as code generation, mathematical reasoning, and healthcare applications, while also identifying areas for future research. As AI continues to evolve, DeepSeek's contributions will play a crucial role in shaping the future of artificial intelligence.

The advancements introduced by DeepSeek mark a significant shift in the landscape of large language models (LLMs), offering a blend of technical innovation, scalability, and accessibility. Its Mixture-of-Experts (MoE) architecture and Multi-Head Latent Attention (MLA) enable efficient parameter utilization, enhancing performance without incurring prohibitive computational costs. These design choices underscore DeepSeek's ability to process expansive datasets while maintaining high accuracy and inference efficiency, as demonstrated by its consistent success across benchmarks like miniF2F and tasks such as diagnostic modeling in healthcare.

DeepSeek's performance across technical and domain-specific applications has positioned it as a competitive alternative to proprietary models like ChatGPT and Claude. For instance, the model excels in formal reasoning, where grammatical precision and structured writing are paramount. In healthcare, DeepSeek supports diagnostic accuracy and patient management workflows, contributing to optimized operations. Similarly, in finance, DeepSeek exhibits promise in predictive analytics

and risk management, although comparative studies suggest areas for improvement in creative and abstract problem-solving.

One of DeepSeek's standout features is its open-source nature, which democratizes AI research by lowering barriers for innovation in resource-constrained settings. This aligns with global trends toward making cutting-edge technologies universally accessible. While its technical documentation and collaborative potential fuel active research communities, further refinement is needed in multimodal integration and applications in edge computing, areas where its competitors currently hold an edge.

Generative AI, particularly models like DeepSeek, holds immense potential for transforming financial risk management. By leveraging advanced techniques such as prompt engineering, data engineering, and agentic frameworks, financial institutions can enhance their risk assessment capabilities and improve market stability. However, addressing the challenges associated with generative AI will be crucial for realizing its full potential in this domain. Despite its advancements, DeepSeek faces challenges related to scalability, ethical considerations, and computational costs [50]. Future work should focus on improving energy efficiency and reducing biases.

References

1. Al-Garaady, J.; Albuhairey, M.M. Understanding User Perceptions of DeepSeek: A Mixed-Methods Sentiment and Thematic Analysis, 2025, [5172367]. <https://doi.org/10.2139/ssrn.5172367>.
2. AlAfnan, M.A. DeepSeek Vs. ChatGPT: A Comparative Evaluation of AI Tools in Composition, Business Writing, and Communication Tasks. *Journal of Artificial Intelligence and Technology* 2025. <https://doi.org/10.37965/jait.2025.0740>.
3. DeepSeek-AI; Bi, X.; Chen, D.; Chen, G.; Chen, S.; Dai, D.; Deng, C.; Ding, H.; Dong, K.; Du, Q.; et al. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism, 2024, [arXiv:cs/2401.02954]. <https://doi.org/10.48550/arXiv.2401.02954>.
4. DeepSeek-AI; Liu, A.; Feng, B.; Wang, B.; Wang, B.; Liu, B.; Zhao, C.; Dengr, C.; Ruan, C.; Dai, D.; et al. DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model, 2024, [arXiv:cs/2405.04434]. <https://doi.org/10.48550/arXiv.2405.04434>.
5. DeepSeek . <http://www.kjdb.org/CN/10.3981/j.issn.1000-7857.2025.02.00183>.
6. Guo, D.; Zhu, Q.; Yang, D.; Xie, Z.; Dong, K.; Zhang, W.; Chen, G.; Bi, X.; Wu, Y.; Li, Y.K.; et al. DeepSeek-Coder: When the Large Language Model Meets Programming – The Rise of Code Intelligence, 2024, [arXiv:cs/2401.14196]. <https://doi.org/10.48550/arXiv.2401.14196>.
7. Hayder, W.A. Highlighting DeepSeek-R1: Architecture, Features and Future Implications. *International Journal of Computer Science and Mobile Computing* 2025, 14, 1–13. <https://doi.org/10.47760/ijcsmc.2025.v14i02.001>.
8. Gao, T.; Jin, J.; Ke, Z.T.; Moryoussef, G. A Comparison of DeepSeek and Other LLMs, 2025, [arXiv:cs/2502.03688]. <https://doi.org/10.48550/arXiv.2502.03688>.
9. Gupta, R. Comparative Analysis of DeepSeek R1, ChatGPT, Gemini, Alibaba, and LLaMA: Performance, Reasoning Capabilities, and Political Bias.
10. Jiang, Q.; Gao, Z.; Karniadakis, G.E. DeepSeek vs. ChatGPT vs. Claude: A Comparative Study for Scientific Computing and Scientific Machine Learning Tasks, 2025, [arXiv:cs/2502.17764]. <https://doi.org/10.48550/arXiv.2502.17764>.
11. Manik, M.M.H. ChatGPT vs. DeepSeek: A Comparative Study on AI-Based Code Generation, 2025, [arXiv:cs/2502.18467]. <https://doi.org/10.48550/arXiv.2502.18467>.
12. Shakya, R.; Vadiie, F.; Khalil, M. A Showdown of ChatGPT vs DeepSeek in Solving Programming Tasks, 2025, [arXiv:cs/2503.13549]. <https://doi.org/10.48550/arXiv.2503.13549>.
13. Fernandes, D.; Matos-Carvalho, J.P.; Fernandes, C.M.; Fachada, N. DeepSeek-V3, GPT-4, Phi-4, and LLaMA-3.3 Generate Correct Code for LoRaWAN-related Engineering Tasks, 2025, [arXiv:cs/2502.14926]. <https://doi.org/10.48550/arXiv.2502.14926>.
14. Huang, D.; Wang, Z. Explainable Sentiment Analysis with DeepSeek-R1: Performance, Efficiency, and Few-Shot Learning, 2025, [arXiv:cs/2503.11655]. <https://doi.org/10.48550/arXiv.2503.11655>.
15. Hussain, Z.S.; Delsoz, M.; Elahi, M.; Jerkins, B.; Kanner, E.; Wright, C.; Munir, W.M.; Soleimani, M.; Djalilian, A.; Lao, P.A.; et al. Performance of DeepSeek, Qwen 2.5 MAX, and ChatGPT Assisting in Diagnosis of Corneal Eye Diseases, Glaucoma, and Neuro-Ophthalmology Diseases Based on Clinical Case Reports, 2025. <https://doi.org/10.1101/2025.03.14.25323836>.

16. Mondillo, G.; Colosimo, S.; Perrotta, A.; Frattolillo, V.; Masino, M. Comparative Evaluation of Advanced AI Reasoning Models in Pediatric Clinical Decision Support: ChatGPT O1 vs. DeepSeek-R1, 2025. <https://doi.org/10.1101/2025.01.27.25321169>.
17. de Paiva, L.F.; Luijten, G.; Puladi, B.; Egger, J. How Does DeepSeek-R1 Perform on USMLE?, 2025. <https://doi.org/10.1101/2025.02.06.25321749>.
18. Habib Lantyer, V. How U.S. Trade Sanctions Fueled Chinese Innovation in AI: The DeepSeek Case, 2025, [5112973]. <https://doi.org/10.2139/ssrn.5112973>.
19. Krause, D. DeepSeek and FinTech: The Democratization of AI and Its Global Implications, 2025, [5116322]. <https://doi.org/10.2139/ssrn.5116322>.
20. Krause, D. DeepSeek's Potential Impact on the Magnificent 7: A Valuation Perspective, 2025, [5117909]. <https://doi.org/10.2139/ssrn.5117909>.
21. Sallam, M.; Al-Mahzoum, K.; Sallam, M.; Mijwil, M.M. DeepSeek: Is It the End of Generative AI Monopoly or the Mark of the Impending Doomsday? *Mesopotamian Journal of Big Data* **2025**, 2025, 26–34. <https://doi.org/10.58496/MJBD/2025/002>.
22. Wu, J. The Rise of DeepSeek: Technology Calls for the “Catfish Effect”. *Journal of Thoracic Disease* **2025**, 17, 1106–1108. <https://doi.org/10.21037/jtd-2025b-02>.
23. Parmar, M.; Govindarajulu, Y. Challenges in Ensuring AI Safety in DeepSeek-R1 Models: The Shortcomings of Reinforcement Learning Strategies, 2025, [arXiv:cs/2501.17030]. <https://doi.org/10.48550/arXiv.2501.17030>.
24. Olson, M.L.; Ratzlaff, N.; Hinck, M.; Luo, M.; Yu, S.; Xue, C.; Lal, V. Semantic Specialization in MoE Appears with Scale: A Study of DeepSeek R1 Expert Specialization, 2025, [arXiv:cs/2502.10928]. <https://doi.org/10.48550/arXiv.2502.10928>.
25. Performance Optimization of DeepSeek MoE Architecture in Multi-Scale Prediction of Stock Returns. *World Journal of Information Technology* **2025**, 3. <https://doi.org/10.61784/wjit3026>.
26. Wang, C.; Kantarcioglu, M. A Review of DeepSeek Models' Key Innovative Techniques, 2025, [arXiv:cs/2503.11486]. <https://doi.org/10.48550/arXiv.2503.11486>.
27. Xin, H.; Ren, Z.Z.; Song, J.; Shao, Z.; Zhao, W.; Wang, H.; Liu, B.; Zhang, L.; Lu, X.; Du, Q.; et al. DeepSeek-Prover-V1.5: Harnessing Proof Assistant Feedback for Reinforcement Learning and Monte-Carlo Tree Search, 2024, [arXiv:cs/2408.08152]. <https://doi.org/10.48550/arXiv.2408.08152>.
28. Lu, H.; Liu, W.; Zhang, B.; Wang, B.; Dong, K.; Liu, B.; Sun, J.; Ren, T.; Li, Z.; Yang, H.; et al. DeepSeek-VL: Towards Real-World Vision-Language Understanding, 2024, [arXiv:cs/2403.05525]. <https://doi.org/10.48550/arXiv.2403.05525>.
29. Piplani, T.; Bamman, D. DeepSeek: Content Based Image Search & Retrieval, 2018, [arXiv:cs/1801.03406]. <https://doi.org/10.48550/arXiv.1801.03406>.
30. Thoughts on the DeepSeek Triggered Path of AI Development. <http://www.kjdb.org/EN/abstract/article/1000-7857/17790>.
31. Thoughts on the DeepSeek Triggered Path of AI Development. <http://www.kjdb.org/EN/10.3981/j.issn.1000-7857.2025.02.00183>.
32. (PDF) Grok, Gemini, ChatGPT and DeepSeek: Comparison and Applications in Conversational Artificial Intelligence. https://www.researchgate.net/publication/389065042_Grok_Gemini_ChatGPT_and_DeepSeek_Comparison_and_Applications_in_Conversational_Artificial_Intelligence.
33. Ziba-Kulawik, K. Generative AI: New Framework of Using Large Language Models for Analysing Descriptive Qualitative Data, 2025. https://doi.org/10.31235/osf.io/rk29m_v1.
34. Puspitasari, F.D.; Zhang, C.; Dam, S.K.; Zhang, M.; Kim, T.H.; Hong, C.S.; Bae, S.H.; Qin, C.; Wei, J.; Wang, G.; et al. DeepSeek Models: A Comprehensive Survey of Methods and Applications.
35. Neha, F.; Bhati, D. A Survey of DeepSeek Models.
36. Piastou, M. Efficiency and safety of the DeepSeek R1 model compared to OpenAI models. *CYRP* **2025**.
37. Maiti, A.; Adewumi, S.; Tikure, T.A.; Wang, Z.; Sengupta, N.; Sukhanova, A.; Jana, A. Comparative Analysis of OpenAI GPT-4o and DeepSeek R1 for Scientific Text Categorization Using Prompt Engineering, 2025, [arXiv:cs/2503.02032]. <https://doi.org/10.48550/arXiv.2503.02032>.
38. DeepSeek-AI.; Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; et al. DeepSeek-V3 Technical Report, 2025, [arXiv:cs/2412.19437]. <https://doi.org/10.48550/arXiv.2412.19437>.
39. Chen, J.; Zhang, Q. DeepSeek Reshaping Healthcare in China's Tertiary Hospitals, 2025, [arXiv:cs/2502.16732]. <https://doi.org/10.48550/arXiv.2502.16732>.

40. Albuhairey, M.M.; Algaraady, J. DeepSeek vs. ChatGPT: Comparative Efficacy in Reasoning for Adults' Second Language Acquisition Analysis. *Irani Studies* **2025**, pp. 864–883. <https://doi.org/10.55074/hesj.vi44.1313>.
41. Chowdhury, M.N.U.R.; Haque, A.; Ahmed, I. DeepSeek vs. ChatGPT: A Comparative Analysis of Performance, Efficiency, and Ethical AI Considerations.
42. Okaiyeto, S.A.; Bai, J.; Wang, J.; Mujumdar, A.S.; Xiao, H. Success of DeepSeek and Potential Benefits of Free Access to AI for Global-Scale Use. *International Journal of Agricultural and Biological Engineering* **2025**, *18*, 304–306. <https://doi.org/10.25165/ijabe.v18i1.9733>.
43. Allen, R. DeepSeek and AI Innovation : How Chinese Universities Broke through the Glass Ceiling of Technological Advancement. *American Journal of STEM Education* **2025**, *6*, 1–10. <https://doi.org/10.32674/10bz2d04>.
44. Arrieta, A.; Ugarte, M.; Valle, P.; Parejo, J.A.; Segura, S. O3-Mini vs DeepSeek-R1: Which One Is Safer?, 2025, [arXiv:cs/2501.18438]. <https://doi.org/10.48550/arXiv.2501.18438>.
45. Aydin, O.; Karaarslan, E.; Erenay, F.S.; Bacanin, N. Generative AI in Academic Writing: A Comparison of DeepSeek, Qwen, ChatGPT, Gemini, Llama, Mistral, and Gemma, 2025, [arXiv:cs/2503.04765]. <https://doi.org/10.48550/arXiv.2503.04765>.
46. Arabiat, O. DeepSeek AI in Accounting: Opportunities and Challenges in Intelligent Automation, 2025, [5116945]. <https://doi.org/10.2139/ssrn.5116945>.
47. Chen, Y.; Shen, J.; Ma, D. DeepSeek's Impact on Thoracic Surgeons' Work Patterns—Past, Present and Future. *Journal of Thoracic Disease* **2025**, *17*, 1114–1117. <https://doi.org/10.21037/jtd-2025b-04>.
48. China's AI Revolution: How DeepSeek Is Changing the Game. | EBSCOhost. <https://openurl.ebsco.com>.
49. Katta, K. Analyzing User Perceptions of Large Language Models (LLMs) on Reddit: Sentiment and Topic Modeling of ChatGPT and DeepSeek Discussions, 2025, [arXiv:cs/2502.18513]. <https://doi.org/10.48550/arXiv.2502.18513>.
50. Chen, J.; Tang, G.; Zhou, G.; Zhu, W. ChatGPT and Deepseek: Can They Predict the Stock Market and Macroeconomy?, 2025, [arXiv:econ/2502.10008]. <https://doi.org/10.48550/arXiv.2502.10008>.
51. Sapkota, R.; Raza, S.; Karkee, M. Comprehensive Analysis of Transparency and Accessibility of ChatGPT, DeepSeek, And Other SoTA Large Language Models, 2025, [arXiv:cs/2502.18505]. <https://doi.org/10.48550/arXiv.2502.18505>.
52. Islam, C.M.; Chacko, S.J.; Horne, P.; Liu, X. DeepSeek on a Trip: Inducing Targeted Visual Hallucinations via Representation Vulnerabilities, 2025, [arXiv:cs/2502.07905]. <https://doi.org/10.48550/arXiv.2502.07905>.
53. Joshi Satyadhar. Enhancing Structured Finance Risk Models (Leland-Toft and Box-Cox) Using GenAI (VAEs GANs). *IJSRA* **2025**, *14*, 1618–1630.
54. Satyadhar, J. Gen AI for Market Risk and Credit Risk [Ebook ISBN: 9798230094388]. *Draft2Digital Publications Ebook ISBN: 9798230094388* **2025**.
55. Joshi Satyadhar. Using Gen AI Agents With GAE and VAE to Enhance Resilience of US Markets. *The International Journal of Computational Science, Information Technology and Control Engineering (IJCSITCE)* **2025**, *12*, 23–38.
56. Joshi, Satyadhar. Leveraging prompt engineering to enhance financial market integrity and risk management. *World Journal of Advanced Research and Reviews WJARR* **2025**, *25*, 1775–1785.
57. Joshi Satyadhar. Quantitative Foundations for Integrating Market, Credit, and Liquidity Risk with Generative AI. <https://www.preprints.org/> **2025**.
58. Satyadhar, J. Review of Gen AI Models for Financial Risk Management. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN : 2456-3307* **2025**, *11*, 709–723.
59. Satyadhar Joshi. The synergy of generative AI and big data for financial risk: Review of recent developments. *IJFMR-International Journal For Multidisciplinary Research* **2025**, *7*.
60. Satyadhar, J. ADVANCING FINANCIAL RISK MODELING: VASICEK FRAMEWORK ENHANCED BY AGENTIC GENERATIVE AI. *International Research Journal of Modernization in Engineering Technology and Science* **2025**, *7*, 4413–4420.
61. Joshi Satyadhar. Implementing gen AI for increasing robustness of US financial and regulatory system. *International Journal of Innovative Research in Engineering and Management* **2024**, *11*, 175–179.
62. Satyadhar Joshi. Agentic Generative AI and the Future US Workforce: Advancing Innovation and National Competitiveness. *International Journal of Research and Review* **2025**, *12*, 102–113.
63. Joshi, .S. A Literature Review of Gen AI Agents in Financial Applications: Models and Implementations. *International Journal of Science and Research (IJSR)* **2025**, *12*, 1094–1100.

64. Satyadhar Joshi . The Transformative Role of Agentic GenAI in Shaping Workforce Development and Education in the US. *Iconic Research And Engineering Journals* **2025**, 8, 199–206.
65. Joshi, S. A Comprehensive Review of Data Pipelines and Streaming for Generative AI Integration: Challenges, Solutions, and Future Directions.
66. Satyadhar, J. Retraining US Workforce in the Age of Agentic Gen AI: Role of Prompt Engineering and Up-Skilling Initiatives. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)* **2025**, 5.
67. Joshi Satyadhar. Generative AI: Mitigating Workforce and Economic Disruptions While Strategizing Policy Responses for Governments and Companies. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) ISSN (Online) 2581-9429* **2025**, 5, 480–486.
68. Satyadhar, J. Training US Workforce for Generative AI Models and Prompt Engineering: ChatGPT, Copilot, and Gemini. *International Journal of Science, Engineering and Technology ISSN (Online): 2348-4098* **2025**, 13.
69. Satyadhar Joshi. Introduction to Vector Databases for Generative AI: Applications, Performance, Future Projections, and Cost Considerations. *International Advanced Research Journal in Science, Engineering and Technology ISSN (O) 2393-8021, ISSN (P) 2394-1588* **2025**, 12, 79–93.
70. Satyadhar Joshi. Bridging the AI Skills Gap: Workforce Training for Financial Services. *International Journal of Innovative Science and Research Technology* **2025**, 10, 1023–1030.
71. Joshi, S. Introduction to Generative AI and DevOps: Synergies, Challenges and Applications.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.