

Efficiency and safety of the DeepSeek R1 model compared to OpenAI models*Piastou Mikita**University of West Georgia**Master's student***Abstract**

The development of advanced Large Language Models (LLMs) has led to a growing importance in the selection of a particular model for use in real-world scenarios, including science, education, and information handling. Consequently, a comparative study of the new DeepSeek R1 model and OpenAI models is relevant. The objective of this paper is to evaluate the performance and security of the DeepSeek R1 model against recent OpenAI models in terms of prediction accuracy, resilience to attacks, and ethics and privacy compliance. The findings demonstrate that DeepSeek R1 can perform better in tasks requiring detailed and sound logical reasoning, outperforming o1 models, but is inferior to o3 models, especially in terms of mathematical ability. The quality of DeepSeek R1 application in various scientific and practical fields is a subject that requires further research.

Keywords: DeepSeek R1, OpenAI, ChatGPT, GPT4.0, o1 and o3-model, LLM, generative AI.

Эффективность и безопасность модели DeepSeek R1 в сравнении с моделями OpenAI*Пестов Никита Владимирович**Университет Западной Джорджии**студент магистратуры***Аннотация**

С развитием способностей передовых больших языковых моделей (LLMs) все более актуальными становятся вопросы выбора той или иной модели для использования в реальных сценариях в науке, образовании, работы с информацией и другим. Потому актуально сравнительное исследование новой модели DeepSeek R1 и моделей OpenAI. Целью данной работы является оценка производительности и безопасности модели DeepSeek R1 в сравнении с последними моделями OpenAI с точки зрения точности прогнозов, устойчивости к атакам, а также соответствия требованиям этики и конфиденциальности. Показано, что DeepSeek R1 может показывать себя лучше в задачах, требующих подробных и обоснованных логических рассуждений, превосходит по качеству модели o1, но уступает моделям o3, особенно по математическим способностям. Выявлено, что качество применения DeepSeek R1 в различных научных и практических областях пока остается предметом будущих научных исследований.

Ключевые слова: DeepSeek R1, OpenAI, ChatGPT, GPT4.0, o1 и o3-модель, LLM, генеративный ИИ.

Introduction

Recent years have seen a rapid development in the field of artificial intelligence (AI), with deep learning models finding applications in all branches of human endeavor. Among the most notable recent advances is DeepSeek R1, a model developed by the Chinese startup DeepSeek. This model has attracted significant attention due to its performance, which is almost equal to that of the OpenAI-o1 model, as well as its free usage model and the ability to run the model locally on one's computer. A notable milestone was achieved when DeepSeek R1 became the most downloaded free application in the US App Store within a short span of just one month after its launch in January 2025, surpassing the popularity of OpenAI's ChatGPT [1]. However, despite the model's noteworthy performance in model tests on various benchmarks and the developers' observation of its cost-effectiveness in development, a pressing question remains: namely, the comparison of the effectiveness of these models in different real-life scenarios, including exact and natural sciences, medicine, industry, and the management of large volumes of specific information.

Various approaches have been actively used in large language model (LLM) comparison studies, but comparisons with DeepSeek R1 are not yet available, as it is recent. The performance of models in mathematics has been the most actively studied [2]. This paper discusses how LLMs can help mathematicians and what their limitations are. Another paper [3] investigates the use of LLMs for cybersecurity. The paper reviews existing models, including GPT and BERT, and analyses their advantages and disadvantages in different aspects of cyber defense. A recent paper [4] is the only one published so far to consider DeepSeek, but in the context of stock market forecasting and macroeconomics. The paper compares ChatGPT and DeepSeek in their ability to analyze news from the Wall Street Journal and predict market trends. According to the study, ChatGPT has a higher predictive ability, especially for positive news.

The objective of this paper is to assess the performance and security of the DeepSeek R1 model in comparison to the most recent OpenAI-o1 and o3 models with regard to prediction accuracy, resilience to attacks, and ethics and privacy compliance.

A little bit about the DeepSeek R1

The distinguishing feature of DeepSeek R1 is its capacity to reason in a manner analogous to OpenAI models, yet with enhanced intricacy, a quality referred to as Visible Chain-of-Thought [5]. This distinction is accomplished through the Mixture-of-Experts Architecture and Group Relative Policy Optimization employed, setting it apart from OpenAI Transformer models. A notable strength of DeepSeek R1 is its cost-effectiveness. The processing cost of 1 million tokens in DeepSeek R1 is approximately \$0.55, whereas OpenAI o1 incurs a cost of \$15 (see Figure 1 for a comparison of the DeepSeek and OpenAI models) [6].

It should be noted that a comparison with OpenAI-o3-mini and o3-mini-high and o3 is not presented here. It is acknowledged that the AIME and MATH-500 benchmarks are no longer optimal for comparing the mathematical abilities of AI models, as there is a FrontierMath benchmark whose tasks the o3-mini model can solve 32% of simple and 28% of complex problems. There is no data on DeepSeek R1 for this benchmark yet, which makes it difficult to assess the suitability of the model to help solve complex abstract mathematical problems.

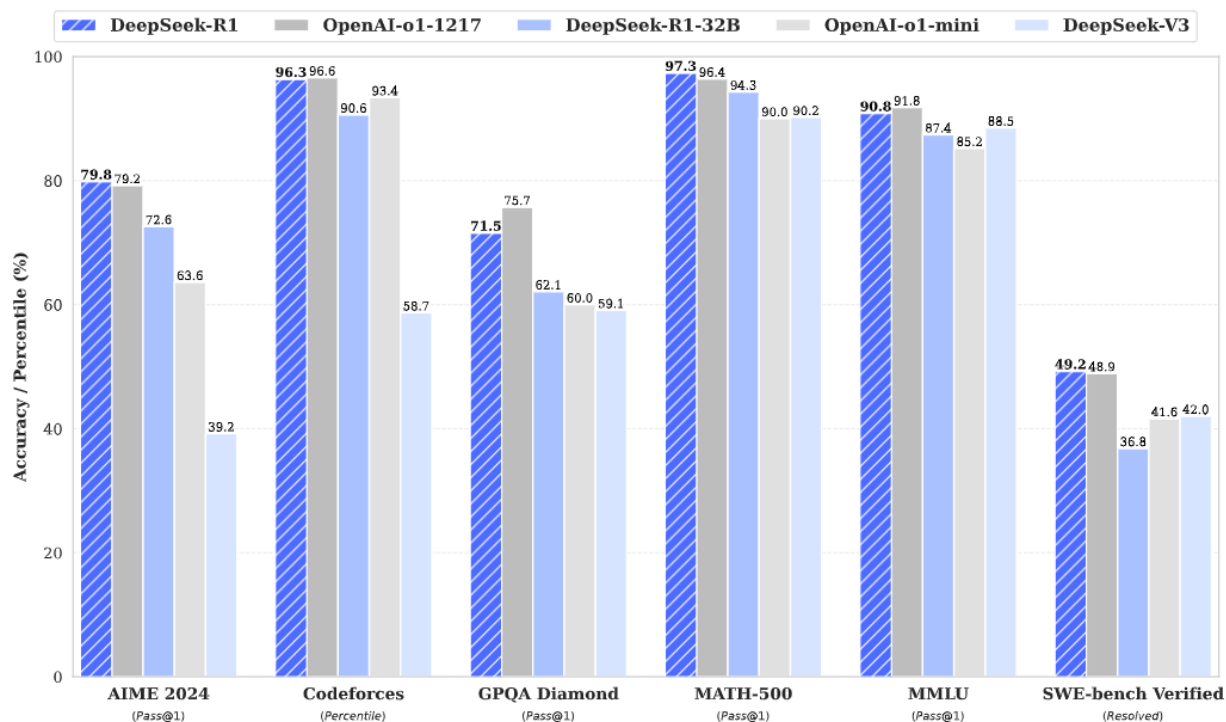


Figure 1 - Comparison of OpenAI and DeepSeek models on some benchmarks

Security and privacy

The present privacy policy of DeepSeek delineates the personal information that the company collects, the manner in which it is used, stored, and shared [7]. The types of data that are typically collected encompass date of birth, username, email address or phone number, password, text or audio inputs, downloaded files, chat history and feedback, device technical data including model, operating system, IP address, unique device identifiers, system language, and service usage information and cookies. Additionally, data from third-party services such as Apple or Google is collected if their authorization is employed, as well as information from advertising and analytics partners. The privacy policy further stipulates that all information collected is stored on secure servers located in the People's Republic of China. It is also noted that DeepSeek may share personal information with its affiliates, service providers and business partners, and where required by law.

Notwithstanding the comprehensive delineation of data protection measures in the privacy policy, concerns have been raised regarding the security and privacy of DeepSeek's utilization of the DeepSeek service. For instance, the storage of user data on servers located in China has been identified as a potential cause for concern,

particularly in countries that adhere to stringent data protection legislation. The Italian data protection regulator, Garante, has mandated the restriction of data processing for Italian users, citing concerns regarding transparency in data processing [8]. Furthermore, research undertaken by Wiz's team revealed that the database was accessible to the public without the requirement of authentication, thus raising concerns over data security and privacy [9]. The database in question contained a substantial amount of data, including user chat history, API secret keys, and internal operational data, amounting to over one million log lines. Moreover, the public accessibility of the database enabled the execution of arbitrary SQL queries, a potential vulnerability that could lead to privilege escalation and the compromise of DeepSeek's internal systems. Following the notification from Wiz, DeepSeek reacted promptly by closing access to the database.

On 6 February 2025, NowSecure published the results of a study which revealed significant security and privacy issues with the DeepSeek iOS app [10]. The report indicated that the app transmits sensitive user data over the internet without adequate protection mechanisms, utilizing outdated and susceptible encryption methods. Specifically, the app disables iOS's built-in App Transport Security (ATS), which is designed to protect data transmitted over the internet. This results in the transmission of information without encryption, thus rendering it susceptible to interception by potential attackers. Additionally, DeepSeek utilizes the outdated 3DES encryption algorithm, a method that has been widely recognised as insecure and susceptible to attack. Modern encryption standards such as AES offer a significantly higher level of security, yet DeepSeek does not employ them.

Finally, it is noteworthy that data leaks and vulnerabilities of a similar nature have been reported in the past in OpenAI as well. A notable distinction between the two platforms is that DeepSeek permits model execution on a local personal computer, a feature not offered by OpenAI, which is known to prioritise user security.

Real-life usage scenarios

A recent research paper analyses the capabilities and limitations of the DeepSeek-R1 model in healthcare [11]. The study addresses the performance of the model, its uses in a medical environment, and the ethical aspects of its application. The authors acknowledge that there are no published studies that have yet tested the capabilities of DeepSeek in analyzing clinical texts and medical images, which leaves open the question of its practical application in medicine. In contrast to the GPT-4o, which furnishes unambiguous responses, the DeepSeek-R1 evinces a propensity for more profound cognitive processing. To illustrate, when posed with a query concerning intraocular metastases in women, DeepSeek acknowledged the absence of a formal categorization of breast cancer as a gynecological disease, yet nevertheless proffered this possibility as the most probable. This manner of response, as posited by the authors, has the potential to enhance clinicians' confidence in the veracity of AI conclusions by rendering the reasoning process more lucid. Nevertheless, in clinical practice, the presence of excessive detail has the potential to impose a significant cognitive burden on clinicians, thereby

impeding the efficiency of the decision-making process. DeepSeek is available in two forms: via application programming interfaces (APIs) and for local deployment, a feature that renders it a versatile instrument for healthcare providers. However, it is important to note that the utilization of APIs can give rise to concerns regarding patient data privacy, as information is transferred to external servers. The local deployment option has been demonstrated to resolve the aforementioned issue by enabling healthcare providers to store and process data within their infrastructure.

To date, there have been no studies conducted on the application of DeepSeek in various fields; however, there are individual expert opinions based on experiments. For instance, physics doctor and neuroscientist Alexander Unzicker has demonstrated the capabilities of DeepSeek R1 and OpenAI models in fundamental physics using examples from electrodynamics, relativity theory, and cosmology [12]. Unzicker demonstrates the utility of DeepSeek R1 in enhancing comprehension of extensive model reasoning chains and in the generation of unanticipated concepts. Examples of logical reasoning experiments encompass the identification and resolution of contradictions, the analysis of causal and correlation relationships, comparative multivariate and contextual analyses, the resolution of ethical dilemmas, and problem planning [13]. The efficacy of DeepSeek in such tasks is demonstrated, though it is acknowledged that the program exhibits deficiencies in the analysis of highly abstract tasks and complex temporal reasoning.

In addition, the response time is a crucial factor in this context. The o3-mini demonstrates enhanced response times in comparison to its predecessors. Specifically, the average response time is recorded as 7.7 seconds, which represents a 24% reduction compared to the o1-mini's response time of 10.16 seconds. The DeepSeek-R1 employs a distinctive approach to query processing, dividing complex problems into stages and analysing them sequentially. Consequently, the algorithm can require several hundred seconds to solve complex questions, for instance, when tackling fundamental physics problems.

Conclusion

DeepSeek R1 is a noteworthy contender in the field of artificial intelligence, particularly when evaluated in terms of cost-effectiveness. Its potential as a serious competitor to the closed OpenAI models is a subject that merits further investigation. The present paper demonstrates the efficacy of DeepSeek R1 in a range of scientific and practical applications, including linguistic analysis, programming and basic sciences. However, the absence of formal data in comparison to new benchmarks, such as FrontierMath, hinders the ability to make a definitive judgement on its suitability in the exact sciences. In the medical domain, DeepSeek R1 has been demonstrated to exhibit a profound comprehension of context, in addition to the capacity to execute meticulous step-by-step analyses. This renders it a potentially valuable instrument in the realm of clinical decision-making. Notwithstanding the assertions made by the developers concerning the safeguarding of data, concerns have been raised regarding the occurrence of information leaks and the utilisation of obsolete encryption methodologies. Conversely, the capacity to execute local deployment of DeepSeek R1 on personal hardware has been identified as a measure

that can be employed to mitigate these risks, a capability that is not presently offered by the closed nature of OpenAI models. In the future, it is imperative to conduct independent testing in real-world use cases and to develop new standards for assessing the quality and security of AI models.

References

1. DeepSeek - AI assistant // App Store. URL: <https://apps.apple.com/ru/app/deepseek-%D0%B8%D0%B8-%D0%B0%D1%81%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BD%D1%82/id6737597349> (date accessed: 27.01.2025).
2. Large Language Models for Mathematicians / S. Frieder, J. Berner, P. Petersen, T. Lukasiewicz // arXiv.2023. URL: <https://arxiv.org/abs/2312.04556> (date accessed: 17.02.2025).
3. Hassanin M. A, Moustafa N. Comprehensive Overview of Large Language Models (LLMs) for Cyber Defences: Opportunities and Directions. A Comprehensive Overview of Large Language Models (LLMs) for Cyber Defences // arXiv. 2024. URL: <https://arxiv.org/abs/2405.14487> (date accessed: 17.02.2025).
4. ChatGPT and Deepseek: Can They Predict the Stock Market and Macroeconomy? ChatGPT and Deepseek / J. Chen, G. Tang, G. Zhou, W. Zhu. // arXiv. 2025. URL: <https://arxiv.org/abs/2502.10008> (date accessed: 17.02.2025).
5. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. DeepSeek-R1 / DeepSeek-AI, D. Guo, D. Yang et al. // arXiv. 2025. URL: <https://arxiv.org/abs/2501.12948> (date accessed: 12.02.2025).
6. OpenAI o3-mini // OpenAI. URL: <https://openai.com/index/openai-o3-mini/> (date accessed: 10.02.2025).
7. DeepSeek Privacy Policy // DeepSeek. URL: <https://chat.deepseek.com/downloads/DeepSeek%20Privacy%20Policy.html> (date accessed: 07.02.2025).
8. Italy's regulator blocks Chinese AI app DeepSeek on data protection / S. Rossi, G. Vagnoni, A. Cawthorne, S. Maler, D. Babington, L. Heavens // Reuters. URL: https://www.reuters.com/technology/artificial-intelligence/italys-privacy-watchdog-blocks-chinese-ai-app-deepseek-2025-01-30/?utm_source=chatgpt.com (date accessed: 08.02.2025).
9. Wiz Research Uncovers Exposed DeepSeek Database Leaking Sensitive Information, Including Chat History. Gal Nagli. // Wiz. URL: <https://www.wiz.io/blog/wiz-research-uncovers-exposed-deepseek-database-leak> (date accessed: 04.02.2025).
10. NowSecure Uncovers Multiple Security and Privacy Flaws in DeepSeek iOS Mobile App. Andrew Hoog. // NowSecure. URL: <https://www.nowsecure.com/blog/2025/02/06/nowsecure-uncovers-multiple-security-and-privacy-flaws-in-deepseek-ios-mobile-app/> (date accessed: 11.02.2025).

11. Peng Y. et al. From GPT to DeepSeek: Significant gaps remains in realizing AI in healthcare //Journal of Biomedical Informatics. 2025. C. 104791. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1532046425000206> (date accessed: 15.02.2025).
12. Unzicker's Real Physics. Alexander Unzicker. // YouTube. URL: <https://www.youtube.com/@TheMachian/videos> (date accessed: 15.02.2025).
13. OpenAI-o3-mini vs DeepSeek R1: Complete Comparison of Advanced AI Reasoning Models Shubham Kumar. // GeeksforGeeks. URL: <https://www.geeksforgeeks.org/openai-o3-mini-vs-deepseek-r1/?ysclid=m74k8qt9bj433476205> (date accessed: 09.02.2025).