ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

Exploration of DeepSeek's Application in Auditing—Based on Tests of Ernie Bot, DeepSeek, Kimi, and Doubao

Jiarui Zeng[1], Ruohan Xia[1], Junqi Zhang[1], Chang Liu[1,*]

[1]*International School, Nanjing Audit University, Nanjing, China*
**Corresponding author*

Abstract: *The DeepSeek large language model (LLM) brings transformative opportunities to the auditing industry with its advantages of high efficiency and low cost, multimodal processing, professional knowledge base, and high-precision analysis. To explore its real-world performance and application prospects in the auditing field, this study conducted a comprehensive evaluation of general-purpose domestic LLMs represented by Ernie Bot, Kimi, and Doubao, as well as the high-performance open-source inference model DeepSeek, using an examination-based testing method across three rounds of quantitative tests. The test data included real questions from the Junior Auditor Qualification Examination. From the results, it is apparent that LLMs such as Ernie Bot, Kimi, and Doubao excel at answering memorization-based questions but have some deficiencies when addressing comprehension-based questions, and thus cannot wholly replace professional auditors. DeepSeek, despite failing to replace experienced auditors, excels in memorization-based questions and has great advantages in answering questions that require critical thinking and deep understanding because of its superior logical capability, fast inference speed, economical training and inference, and outstanding code generating capability. In light of the results presented in this research, we propose an approach for incorporating DeepSeek into auditing practices to take advantage of its capabilities, relieve auditors from excessive workload, improve audit quality and efficiency, and offer valuable insights into the auditing sector in the age of AI.*

Keywords: *DeepSeek, Artificial Intelligence, Smart Auditing*

1. Introduction

In view of the rapid advancement of artificial intelligence technology, Generative Artificial Intelligence (GenAI) has become increasingly common in many sectors, particularly in LLMs. Regarding auditing, DeepSeek has opened up new possibilities and opportunities for transforming the conventional audit work model [1].

DeepSeek is a product of Beijing DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd., designed using the Transformer framework. The DeepSeek large model exhibits extraordinary natural language processing capabilities, capable of handling numerous complex language tasks. The strengths of this model lie in its cost-effectiveness and openness. The model offers impressive performance coupled with efficient cost management while requiring minimal computational resources. Its openness makes it available to external developers who can reuse the model, and businesses can utilize the model on their servers, thus guaranteeing the autonomy and security of data [2]. Due to its exceptional performance and diverse applications, DeepSeek has attracted significant attention from the industry. Many companies and research organizations have found the technology appealing due to its superior performance and versatility. It has sparked intense debate in the field of financial accounting, serving as a catalyst in the adoption of AI technology in the industry [3].

But auditing requires accurate examination of the financial statement, strict evaluation of internal control, and effective risk assessment, which require professional knowledge and rich practical experience of auditors. Furthermore, to meet legal and regulatory requirements, auditors need to be competent in complicated regulations. To face enterprises of various industries and sizes, the auditor also needs adaptability and fast learning ability[4]. As such, very stringent requirements exist for models in conducting audits. The extent to which LLMs can take over the position of experienced auditors and ways to make better use of such models are very urgent topics to explore [5].

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

When asked about challenges and changes brought by DeepSeek to auditing, DeepSeek responds as Figure 1.

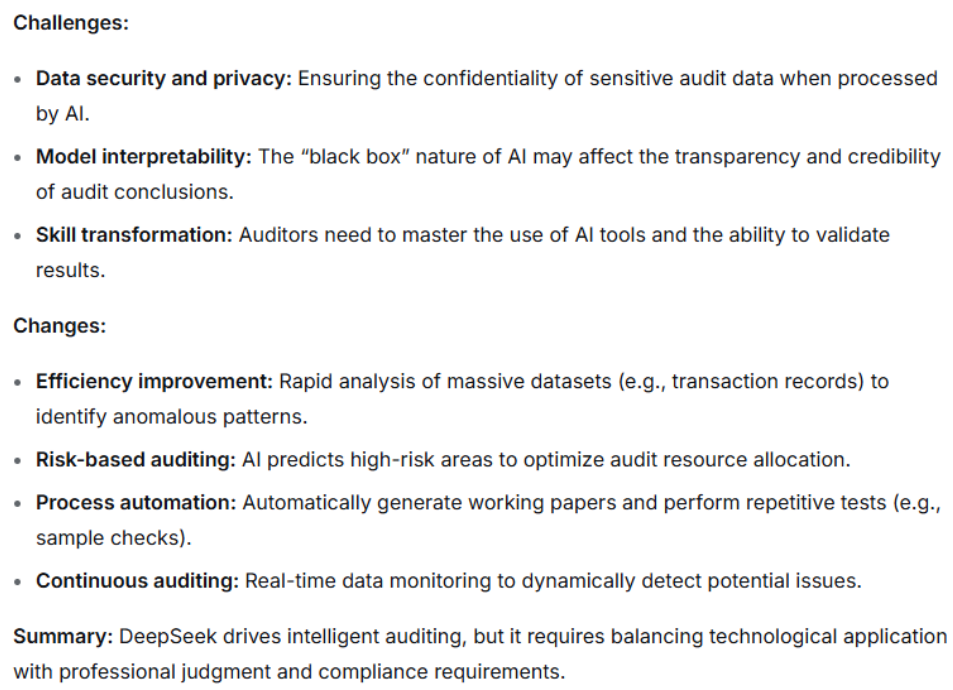

Figure 1: Response diagram of DeepSeek to the challenges and changes in audit work

2. Audit knowledge testing

To test the actual capabilities of Ernie Bot, DeepSeek, Kimi, and Doubao in auditing, to examine their feasibility of use in auditing tasks, and to develop relevant application methods, this study conducts quantitative testing of the four large language models using a paper test approach.

2.1 Test design

To obtain objective and comparable performance data of Ernie Bot, DeepSeek, Kimi, and Doubao in the auditing field, this study designed three rounds of tests using three representative audit examination papers. Test results were calculated and statistically analyzed and compared with human candidates' scores to draw conclusions.

2.1.1 Selection of test questions

Across three rounds of tests, the study selected three sets of real questions from the Junior Auditor Qualification Examination published by Yiwen Education. As authoritative examinations in the auditing field, *Audit Theory and Practice* and *Basic Audit Knowledge* papers are highly standardized and professional, comprehensively assessing candidates' comprehensive application of audit knowledge. Considering the database update timelines of each LLM, the final selected questions are the third, fourth, and fifth sets (recalled by candidates) of *Audit Theory and Practice* and *Basic Audit Knowledge* for the Junior Auditor Qualification Examination.

2.1.2 Test procedures

First, test questions were input sequentially into the dialogue interface of each LLM to complete three rounds of testing. The responses received from the LLMs were recorded to guarantee that the data was not manipulated. After that, calculations and statistical analysis of the test results were performed from the macroscopic and microscopic standpoints, whereby macroscopic analysis concentrated on overall performance, while microscopic analysis was concerned with the sub-modules like audit risk assessment, evidence gathering, and audit report writing. Conclusions were subsequently derived from data analysis.

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

2.2 Test results of recalled junior auditor examination questions

2.2.1 Basic introduction to recalled examination questions

The third and fourth sets of recalled *Audit Theory and Practice* questions and the third and fifth sets of recalled *Basic Audit Knowledge* questions each consist of three question types, totaling 65 questions and 100 points: 30 single-choice questions (30 points), 20 multiple-choice questions (40 points, no points for wrong choices, 0.5 points per correct choice for partial answers), and 15 comprehensive analysis questions (30 points, three cases covering financial management, accounting, and law, each with five multiple-choice questions with 1–4 correct options; no points for wrong choices, 0.5 points per correct choice for partial answers).

Notably, the fifth set of recalled *Audit Theory and Practice* questions totals 62 questions and 95 points: 29 single-choice questions (29 points), 18 multiple-choice questions (36 points, same scoring rules), and 15 comprehensive analysis questions (30 points, same scoring rules). The fourth set of recalled *Basic Audit Knowledge* questions totals 64 questions and 98 points: 30 single-choice questions (30 points), 19 multiple-choice questions (38 points, same scoring rules), and 15 comprehensive analysis questions (30 points, same scoring rules).

2.2.2 Overview of test 1

As shown in Table 1, in the multiple choice questions test, ERNIE Bot got a score of 28 marks with a correctness of 93.33%; DeepSeek and Kimi obtained perfect scores of 30 marks with a correctness of 100%, while Doubao got a score of 29 with a correctness of 96.67%. For the open-ended questions, ERNIE Bot scored 35 points with a correct rate of 87.5%; both DeepSeek and Kimi received full scores of 40 points with a correct rate of 100%; and Doubao scored 40 points with a correct rate of 100%. In terms of comprehensive analysis questions, ERNIE Bot scored 12.5 points with a correct rate of 41.67%; DeepSeek scored 28 points with a correct rate of 93.33%; Kimi achieved full scores of 30 points with a correct rate of 100%; and Doubao scored 26.5 points with a correct rate of 88.33%. Overall, DeepSeek and Kimi performed exceptionally well in both multiple-choice and open-ended questions, achieving full scores. Among them, Kimi performed the best in comprehensive analysis questions, achieving a full score.

Table 1: The third sets of recalled audit theory and practice questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	28	93.33%	30	100%	30	100%	29	96.67%
20 Multiple-Choice Questions (40 points)	35	87.5%	40	100%	40	100%	40	100%
15 Comprehensive Analysis Questions (30 points)	12.5	41.67%	28	93.33%	30	100%	26.5	88.33%

As shown in Table 2, in the multiple-choice section, ERNIE Bot, DeepSeek, Kimi, and Doubao all achieved a perfect score of 30 points, with a 100% accuracy rate. For the multiple-choice questions, ERNIE Bot, DeepSeek, Kimi, and Doubao also scored a full 40 points, maintaining a 100% accuracy rate. In terms of comprehensive analysis questions, ERNIE Bot, DeepSeek, Kimi, and Doubao all secured a perfect score of 30 points, with a 100% accuracy rate. Overall, these four models performed exceptionally well in all three types of questions in this exam, achieving full marks.

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

Table 2: The third sets of recalled basic audit knowledge questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	30	100%	30	100%	30	100%	30	100%
20 Multiple-Choice Questions (40 points)	40	100%	40	100%	40	100%	40	100%
15 Comprehensive Analysis Questions (30 points)	30	100%	30	100%	30	100%	30	100%

2.2.3 Overview of test 2

As shown in Table 3, in the multiple-choice section, ERNIE Bot scored 29 points with a correct rate of 96.67%; DeepSeek, Kimi, and Doubao all achieved a perfect score of 30 points with a correct rate of 100%. For the open-ended questions, ERNIE Bot scored 27.5 points with a correct rate of 68.75%; DeepSeek, Kimi, and Doubao all received a perfect score of 40 points with a correct rate of 100%. In terms of comprehensive analysis questions, ERNIE Bot scored 16.5 points with a correct rate of 100%; DeepSeek and Kimi both achieved a perfect score of 30 points with a correct rate of 100%; Doubao scored 27 points with a correct rate of 90%. Overall, DeepSeek and Kimi performed well in both multiple-choice and open-ended questions, achieving perfect scores. In comprehensive analysis questions, DeepSeek and Kimi also achieved perfect scores, while Doubao's score and correct rate were relatively high. ERNIE Bot performed well in multiple-choice and comprehensive analysis questions, but its score and correct rate in multiple-choice questions were significantly lower than those of other models.

Table 3: The forth sets of recalled audit theory and practice questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	29	96.67%	30	100%	30	100%	30	100%
20 Multiple-Choice Questions (40 points)	27.5	68.75%	40	100%	40	100%	40	100%
15 Comprehensive Analysis Questions (30 points)	16.5	100%	30	100%	30	100%	27	90%

As shown in Table 4, in the multiple-choice section, ERNIE Bot, DeepSeek, Kimi, and Doubao all achieved a perfect score of 30 points, with a 100% accuracy rate. For the multiple-choice questions, ERNIE Bot, DeepSeek, Kimi, and Doubao also scored a full 40 points, with a 100% accuracy rate. In terms of comprehensive analysis questions, ERNIE Bot, DeepSeek, Kimi, and Doubao all secured a perfect score of 30 points, with a 100% accuracy rate. Overall, these four models performed exceptionally well in all three types of questions in this exam, achieving full scores. This indicates that they have a solid grasp of audit-related basic knowledge and are capable of accurately answering various types of

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

questions.

Table 4: The forth sets of recalled basic audit knowledge questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	30	100%	30	100%	30	100%	30	100%
19 Multiple-Choice Questions (38 points)	40	100%	40	100%	40	100%	40	100%
15 Comprehensive Analysis Questions (30 points)	30	100%	30	100%	30	100%	30	100%

2.2.4 Overview of test 3

Regarding the multiple choice questions, as shown in Table 5, ERNIE Bot, DeepSeek, Kimi, and Doubao got full marks and accuracy rates of 100%. It means that all four models have sufficient competence in comprehending and analyzing multiple choice questions. Regarding the open-ended questions, ERNIE Bot got 36 points and an accuracy rate of 94.74%; DeepSeek got 38 points and an accuracy rate of 100%; Kimi got 40 points and an accuracy rate of 100%; and Doubao got 36 points and an accuracy rate of 94.74%. This means that DeepSeek and Kimi have better abilities of comprehending open-ended questions and are capable of giving the right answer among various choices due to their reasoning ability. Regarding the comprehensive analysis questions, ERNIE Bot obtained 27 points and an accuracy rate of 90%; DeepSeek obtained 30 points and an accuracy rate of 100%; Kimi obtained 30 points and an accuracy rate of 100%; and Doubao obtained 27.5 points and an accuracy rate of 91.67%. This implies that DeepSeek and Kimi have better capabilities in dealing with comprehensive analysis questions as they are competent enough to give a deep analysis. ERNIE Bot and Doubao perform well in multiple-choice questions, but their accuracy rates in open-ended and comprehensive analysis questions are slightly lower than those of other models.

Table 5: The fifth sets of recalled audit theory and practice questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	29	100%	29	100%	29	100%	29	100%
19 Multiple-Choice Questions (38 points)	36	94.74%	38	100%	40	100%	36	94.74%
15 Comprehensive Analysis Questions (30 points)	27	90%	30	100%	30	100%	27.5	91.67%

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

In the multiple-choice section, as shown in Table 6, ERNIE Bot, DeepSeek, Kimi, and Doubao all achieved a perfect score of 30, with a 100% accuracy rate. As far as the open questions are concerned, ERNIE Bot, DeepSeek, Kimi, and Doubao managed to achieve a perfect score of 40, with an accuracy rate of 100%. As for the comprehensive analysis questions, ERNIE Bot and Doubao achieved a score of 30 with an accuracy rate of 100%, DeepSeek achieved a score of 28 with an accuracy rate of 100%, and Kimi achieved a score of 30 with an accuracy rate of 100%.

Table 6: The fifth sets of recalled basic audit knowledge questions test record sheet

Question Type	Ernie Bot		DeepSeek		Kimi		Doubao	
	Score	Accuracy	Score	Accuracy	Score	Accuracy	Score	Accuracy
30 Single-Choice Questions (30 points)	30	100%	30	100%	30	100%	30	100%
19 Multiple-Choice Questions (38 points)	40	100%	40	100%	40	100%	40	100%
15 Comprehensive Analysis Questions (30 points)	30	100%	28	100%	30	100%	30	100%

2.3 Error analysis and model capability comparison

2.3.1 Error type distribution

All four models achieved nearly full marks in Basic Audit Knowledge, but significant differentiation appeared in Audit Theory and Practice tests. As shown in Table 7, the average accuracy rates of DeepSeek and Kimi were close to 100%, Doubao 95.8%, and Ernie Bot 81.2%, with differences concentrated in multiple-choice and case analysis questions. In multiple-choice questions, the wrong and missing selection rates of DeepSeek and Kimi were 0%, Doubao about 5.2%, and Ernie Bot as high as 20.7%. Ernie Bot frequently missed correct answers, reflecting its shortcomings in "parallel judgment", that is, the ability to evaluate multiple conditions one by one. In case analysis questions, the average accuracy rates of DeepSeek and Kimi were 97.8% and 100% respectively, Doubao 90%, and Ernie Bot only 62.2%, and Ernie Bot's score fluctuation was as high as 48.3 percentage points, with the worst stability.

Table 7: Comparison of model capability characteristics

Comparative Dimension	DeepSeek	Kimi	Doubao	Ernie Bot
Mastering basic knowledge (Accuracy rate of Basic Audit Knowledge Questions)	100%	100%	100%	100%
Accuracy Assessment of Audit Standards and Procedures(Accuracy rate of Audit Theory and Practice Questions)	99.5%	100%	95.8%	81.2%
Complex Situational Reasoning Ability (Average accuracy rate of case analysis	97.8%	100%	90%	62.2%

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

questions)				
Information Integration and Comprehensive Judgment (Accuracy rate of multiple-choice questions)	100%	100%	94.8%	79.3%
Stability Judgment (Standard Deviation of Scores for Three Sets of Case Analysis Questions)	3.8	0	5.1	24.6
Level of Detailed Analysis	Detailed	Concise	Medium	Medium
Sensitivity to Controversial Topics	May adhere to principles	Not exposed	Medium	Not exposed

2.3.2 Model capability profiles

DeepSeek: Only one discrepancy with the standard answer across all tests (accepting "sequential counting" as a viable but non-optimal practice for inventory supervision). Zero omission rate in case questions makes it suitable for exhaustive internal control defect checking and complete collection of multi-condition evidence, competent for complex reasoning tasks such as risk assessment and audit procedure decision-making. Remark: It might make strict judgments on several acceptable options in reality, necessitating auditor intervention.

Kimi: Perfect performance in all tests, 100% case accuracy, 0% omission, and excellent stability. Very coherent with the standards, dependable for legal provision search and fast audit judgment comparison with standard responses. Compact output allows direct inclusion in working papers; cross-analysis with detailed outputs from other models is advisable for complicated circumstances needing extensive reasoning.

Doubao: Full marks in basic knowledge test, 90.0% case accuracy, 5.2% omission, and decent stability. Points are deducted mainly in complex scenario test questions where there is need for synthesis of standards, procedures, and controls. Reliable for single dimension knowledge searching and classification of data, yet for complicated multiple condition reasoning and risk assessment, it is wise to correlate with DeepSeek/Kimi.

Ernie Bot: Earns 100 percent for the basic knowledge assessment, but only scores 62.2 percent for the accuracy of the case, having an omission rate of 20.7 percent, and is noted as having the lowest stability rating amongst the various software tools. It often omits the correct answers in "select all that apply" question types and lacks understanding of the control points in the business cycle.

2.4 Practical application recommendations

Firstly, at the basic knowledge level, all four models performed excellently. In the "Basic Knowledge Related to Auditing" quiz conducted in six batches, DeepSeek, Kimi, Doubao, and ERNIE Bot all obtained perfect or near-perfect marks, suggesting that the currently popular large models have mastered standardized knowledge retrieval, calculation skills, and regulation answers in the area of auditing.

Secondly, at the level of professional judgment in auditing, significant differentiation emerges among models. DeepSeek and Kimi maintain a 100% accuracy rate in all question types of "Auditing Theory and Practice", demonstrating strong logical reasoning abilities. Doubao achieves an accuracy rate of about 90-92% in case analysis questions, with occasional deviations. ERNIE Bot exhibits significant fluctuations in performance on multiple-choice questions and case analysis questions, especially when multiple correct options need to be comprehensively selected or complex situational reasoning is required, where its accuracy rate is significantly lower than that of other models.

Thirdly, the capability boundaries of various models differ, necessitating differentiated usage. Based on the analysis of incorrect question content, the following application suggestions are proposed. As shown in Table 8, different audit tasks correspond to the recommended models and notes as follows:

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

Table 8: Application recommendations for the four models

Audit Task Type	Recommended Model(s)	Notes
Knowledge retrieval, regulatory inquiry, standardized Q&A	Any model	Manual verification is recommended for the latest regulations
Defect identification in audit working papers	DeepSeek	Strongest capability in handling negative instructions and logical reasoning
Summary generation for long contracts and reports	Kimi or DeepSeek	Kimi has an advantage in ultra-long text processing
Answer verification for multiple-choice questions	DeepSeek or Kimi	Ernie Bot carries a high risk of omission and is not recommended for standalone use
Complex case analysis	DeepSeek or Kimi	Doubao requires manual review; Ernie Bot is not recommended
First-draft generation for standardized documents	Any model	All model-generated content requires manual review
Judgment on ambiguous issues	Do not rely on a single model	Results from multiple models should be integrated, with final decisions made by auditors

3. Exploration of DeepSeek's application in auditing

In the wave of technological empowerment in the 21st century, the deep integration of audit technology and informatization has become an irreversible trend. However, the traditional audit model exposes bottlenecks in digital transformation: difficulties in integrating multi-source heterogeneous data, low process standardization, and redundant resource input. This technological lag restricts risk identification and decision-support capabilities in complex economic environments. Thus, accelerating the construction of a smart audit system is an inevitable choice for industry development and a strategic requirement to improve national audit governance efficiency [6].

When asked about its applications in auditing, DeepSeek lists: data analysis and anomaly detection; risk assessment and prioritization; automated document processing; automated internal control testing; audit evidence collection and organization; intelligent report generation; continuous auditing and real-time monitoring; compliance checking and regulatory updates; knowledge management and collaboration support; fraud investigation and pattern recognition. It also notes challenges: data security (encrypted transmission/storage compliant with GDPR and other regulations); interpretability (providing decision basis such as feature importance analysis to meet working paper requirements); human-AI collaboration (AI handling repetitive tasks, auditors focusing on professional judgment and complex decisions).

3.1 Batch interpretation and summarization of unstructured data

Auditors frequently process large volumes of unstructured data (contracts, meeting minutes, reports, etc.) that are time-consuming and error-prone. DeepSeek's powerful NLU enables batch interpretation and summarization of such texts. Auditors can input text files (.txt, .doc, .pdf, etc.) for semantic analysis, key information extraction, outline generation, and concise summarization, greatly improving efficiency and reducing manual reading costs [7–9].

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

3.2 Intelligent classification of text fields

Facing hundreds of thousands of journal entry descriptions or bank statements, keyword search is often ineffective. DeepSeek can perform intelligent classification of textual audit data after cleaning, labeling, and feature extraction, helping auditors quickly identify abnormal or high-risk transactions and improve accuracy [10].

3.3 Intelligent audit knowledge Q&A

DeepSeek can build an intelligent Q&A system integrated with an audit knowledge base and retrieval-augmented generation (RAG), providing auditors with quick access to regulations, policies, and cases to support on-site decision-making [11–12].

3.4 Audit document generation

Leveraging strong text generation capabilities, DeepSeek automates drafting of audit reports, working papers, and other documents based on templates and input data (sampling results, inventory differences, etc.), assisting in drafting qualified opinions or emphasis-of-matter paragraphs and reducing manual drafting workload [13].

4. Advantages of DeepSeek in auditing

4.1 Improving audit efficiency

The powerful processing ability of DeepSeek can quickly analyze a large amount of data, automatically generate summaries and reports, and greatly shorten the audit cycle. For example, in unstructured data processing, its batch interpretation and summary function enables auditors to avoid reading a large amount of text word by word, save time costs, and devote more energy to in-depth analysis of key issues.

4.2 Enhancing audit quality

Its accurate text understanding and classification ability help auditors identify abnormal situations more accurately [14]. In the scenario of intelligent classification of text fields, DeepSeek can carefully distinguish different types of audit items, reduce manual classification errors, and improve the reliability of audit results. At the same time, its intelligent Q&A system for audit knowledge can provide auditors with accurate regulatory interpretation and case references, helping auditors make more appropriate audit judgments.

4.3 Promoting audit innovation

DeepSeek has promoted the transformation of auditing work from traditional experience-driven to intelligent data-driven, and spawned new audit techniques and methods [15]. For example, using its data mining and analysis functions, auditors can discover potential risks and trends hidden behind complex data, expand audit perspectives, and provide more forward-looking support for audit decision-making.

4.4 Reducing audit costs

DeepSeek can realize the intellectualization of audit processes, reduce the dependence on human resources, and lower labor costs [16]. At the same time, its efficient document processing and generation ability reduce the time investment of auditors in document writing and sorting, improve work efficiency, and indirectly reduce audit costs [17]. In addition, different from the annual fee of traditional SaaS systems or the high API call fee of closed-source large models, the open-source feature of DeepSeek allows low-cost local deployment. This not only saves huge software procurement costs, but also completely solves the security concerns of audit data going abroad or being stored in the cloud.

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

5. Challenges of DeepSeek in auditing

5.1 Data security and privacy protection

Auditing work involves a large amount of sensitive information, such as customer financial data, internal audit reports, etc. When using DeepSeek, it is necessary to ensure the security of data in the process of collection, storage, processing and transmission, and prevent data leakage or improper use [18]. At the same time, it is necessary to abide by relevant laws and regulations, such as the *Data Security Law* and the *Personal Information Protection Law*, to protect the legitimate rights and interests of data subjects.

5.2 Model accuracy and reliability

Although DeepSeek has strong language understanding and generation ability, its output results may still have errors or inaccuracies, the so-called "hallucination" problem, which is extremely serious. It may fabricate a non-existent "Audit Standard No. XX" clause in a perfect audit report draft, or incorrectly splice financial data of different enterprises. In auditing work, the accuracy of information is extremely high, so it is necessary to strictly review the output of DeepSeek to ensure its reliability [19]. Besides, the precision of the model depends upon the quality of the training data and the extent to which the algorithm is optimized, and it becomes essential to constantly work on optimizing the model to enhance its precision.

5.3 Difficulty in technical application and integration

The implementation of DeepSeek technology in the process of auditing entails the need for its successful integration into current audit processes and information systems. It includes both technical and personnel adaptations to the use of new technology. The auditors must have the necessary technical proficiency, effectively apply and utilize DeepSeek technology, and derive the maximum benefit from it [20].

5.4 Regulatory and compliance issues

The use of AI technology in the auditing industry is currently in a phase where it is continuously improving and standardizing, and the regulation policies and standards are constantly being updated [21]. In the process of applying DeepSeek to auditing, one must be very careful about keeping track of the regulation trends to ensure that the application does not violate any regulation standards. Meanwhile, auditing activities themselves also need to follow very rigorous audit standards and professional ethics.

6. Conclusion and outlook

DeepSeek shows tremendous potential for application in auditing, with high efficiency, effectiveness, innovativeness, and cost-efficiency. Yet, DeepSeek encounters numerous obstacles in terms of data safety, precision, integration, and compliance. The application of DeepSeek in auditing will become increasingly common with advancements in technology and improvements in regulation. Auditors need to develop their expertise and skills to fully exploit the advantages of DeepSeek, overcome obstacles, and lead the auditing sector towards intelligence and digitization.

Acknowledgements

This work was funded by the Provincial College Student Innovation and Entrepreneurship Training Program of Nanjing Audit University (Grant No.:S202511287015).

References

[1] Chen W. Research on the application of large models in smart auditing[J]. The Chinese Certified Public Accountant, 2024(05):86–90+5.
[2] Bai Z Y. Challenges, opportunities and countermeasures of DeepSeek for China's accounting industry[J]. Friends of Accounting, 2025(8):143–148.
[3] Zhang X Q. DeepSeek may become a catalyst for the accounting and finance industry to embrace AI

ISSN 2616-7433 Vol. 8, Issue 2: 24-34, DOI: 10.25236/FSST.2026.080204

[N]. China Accounting News, 2025-02-21.
[4] Zhang L, Wei H Q. AI empowers internal audit: deep integration and innovative development[J]. Internal Auditing in China, 2025(04):36–39.
[5] Li X Y. Research on internal audit empowering new quality productivity from the perspective of digital transformation[J]. Modern Auditing and Accounting, 2025(02):14–17.
[6] Wang Y J. Thoughts on innovative paths for smart transformation of audit work in the digital context[J]. Journal of Chifeng University (Natural Science Edition), 2024,40(12):63–66.
[7] Zang A, Chen W, Fan W, et al. AI reshaping auditing: exploring the industry transformation driven by artificial intelligence[J]. The Chinese Certified Public Accountant, 2025(02):70–72.
[8] Bi X F, Zhou S J. Research on smart auditing: thinking transformation and platform construction[J]. Time-Honored Brand Marketing, 2024(19):82–84.
[9] Tian T, Liang Y. Application of large language models in national auditing[J]. Auditing Research, 2025(01):33–41.
[10] Cao Z W, Zhao L, Zhang Q C. Localization exploration of AI technology and its application in national auditing[J]. Auditing Research, 2024(05):26–37.
[11] Xu C, Chen Y, Ge H M, et al. Research on audit technology based on big data[J]. Acta Electronica Sinica, 2020,48(05):1003–1017.
[12] Huang X Z, Lyv Q. Ideas and practical exploration for the construction of an AI audit system[J]. Auditing Research, 2024(06):30–38.
[13] Huang J J, Zhou L Y, Xu C. Research on the construction of an intelligent audit knowledge Q&A system based on large models[J/OL]. Friends of Accounting, 2025(09):24–30.
[14] Wu S J. Application of generative AI in audit risk assessment and risk response[J]. China-Arab Science and Technology Forum, 2025(02):31–35.
[15] Xu J L, Zhang Y. Research on the impact mechanism of internal audit innovation on the development of new quality productivity: a case study of China Merchants Group[J]. Modern Business Trade Industry, 2025(07):181–183.
[16] Li A Q, Wang W Q, Li B Q. Research on the role and mechanism of new quality productivity empowering audit work innovation and development[J]. Enterprise Reform and Management, 2025(04):112–114.
[17] Li X C, Huang Z W, Yang W S, et al. New Quality Productivity Empowers Digital-Intelligent Transformation of Auditing to Promote High-Quality Development of State-Owned Enterprises: Practice of Audit Reform in China Mobile Yunnan Company[C]//China Enterprise Reform and Development Association. 2024 Outstanding Achievements in China Enterprise Reform and Development (8th Session) Volume 1. China Mobile Communications Group Yunnan Co., Ltd., 2025:343–358.
[18] Yu H. Challenges and countermeasures of audit ethics in the digital context[J]. Journal of Chongqing Three Gorges University, 2024,40(05):13–21.
[19] Wu J. Technical accessibility and social governance challenges under DeepSeek application: based on large-scale deployment of generative AI[J/OL]. Journal of Soochow University (Philosophy and Social Science Edition), 1–9.
[20] Wang J J. Research on the theoretical logic and innovative path of new quality productivity empowering high-quality development of internal audit[J]. China Financial Market, 2025(06):165–167.
[21] Zhang Y. Ethical dilemmas and innovative strategies of AI auditing[J]. Co-Operative Economy & Science, 2025(04):155–157.