OPEN ACCESS
Renewable Energy LETTER

Exploring the applications of large language models in carbon management and assessment

Dashu Li,[1,*] Mengdi Liu,[2,*] Jiayi Li,[1] and Jiashuo Cheng[1]
[1]Beijing Key Laboratory of Process Fluid Filtration and Separation, College of Mechanical and Transportation Engineering, China University of Petroleum (Beijing), Beijing 102249, China
[2]School of Management Science and Engineering, Shandong Technology and Business University, Yantai 264005, China
*Correspondence: ls01ax@163.com (D.L.); 202414306@sdtbu.edu.cn (M.L.)
Received: August 20, 2025; Accepted: December 16, 2025; Published Online: January 20, 2026; https://doi.org/10.59717/j.xinn-energy.2026.100135

Citation: Li D., Liu M., Li J., et al. (2026). Exploring the applications of large language models in carbon management and assessment. *The Innovation Energy* **3**:100135.

Dear Editor,

Climate change stands as the most pressing global challenge of the twenty-first century, with the Paris Agreement's carbon neutrality target placing heightened demands on effective carbon footprint management. Large language models (LLMs) have recently emerged as transformative tools in this domain, offering powerful natural language processing capabilities and cross-domain scalability. This research systematically evaluates the performance of mainstream LLMs, including ChatGPT and DeepSeek-R1, identifying significant gaps in their knowledge integration depth, logical reasoning accuracy, and dynamic modeling capabilities. Despite these limitations, the study highlights their vast potential for future advancements. To enhance the applicability of LLMs in carbon management, this paper proposes five key areas for improvement: optimizing knowledge base integration, strengthening logical reasoning capabilities, increasing the efficiency of tool and strategy generation, improving policy adaptability, and advancing dynamic modeling capacities. As these areas are addressed, large language models are poised to become essential intelligence tools in the global effort to achieve carbon neutrality.

LEVERAGING LLMS FOR CARBON FOOTPRINT MANAGEMENT

Climate change is one of the most urgent global challenges of the 21st century. According to the Sixth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC), exceeding 1.5°C of global warming will lead to irreversible ecological damage.[1] The "carbon neutrality" goal outlined in the Paris Agreement has become a key driver for global climate action, prompting countries to gradually adopt mandatory carbon disclosure mechanisms. Consequently, carbon footprint management has emerged as a vital component of sustainable development strategies. However, the complexity of carbon emission accounting and the evaluation of emission reduction technologies pose significant challenges for information extraction, analysis, and decision-making.[2] For instance, carbon emission data often require intricate processes such as supply chain traceability and energy conversion coefficient calculations. Similarly, assessing emission reduction technologies involves integrating multidisciplinary models, including life cycle analysis and marginal abatement cost curves.[3]

Traditional approaches face three primary challenges in this context. First, rule-based systems struggle to adapt to the constantly evolving Monitoring, Reporting, and Verification frameworks of different countries.[4] Second, statistical models lack the capacity to identify implicit policy biases, such as the trade protectionist tendencies embedded in carbon tariffs. Third, the development of expert systems is time-intensive and technically demanding, making it particularly difficult for smaller economies to meet the diverse and complex requirements of carbon management. These challenges highlight the dynamic, multidisciplinary, and policy-driven nature of carbon footprint management, which demands more advanced information processing capabilities.[5]

In recent years, generative artificial intelligence (AI), particularly large language models (LLMs), has opened new pathways to address these challenges. LLMs excel in processing vast amounts of unstructured data and exhibit strong cross-domain adaptability.[6] For example, OpenAI's ChatGPT has gained recognition for its ability to handle technical, scientific, and policy-related texts, while DeepSeek-R1, developed by China's DeepSeek Company, has been noted for its cost-efficient training system and adaptability to multidisciplinary fields, including environmental science. However, the complexity of carbon footprint management imposes higher demands on LLMs. For instance, carbon policy texts often include regional exemptions, such as the transitional provisions in the EU's Carbon Border Adjustment Mechanism. Similarly, emission reduction technology assessments frequently involve complex mathematical models, such as thermal-electric coupling equations.[7] These unique challenges necessitate a closer examination of the current capabilities and limitations of LLMs in carbon footprint management.

Based on this, this study employs a multi-dimensional test set and assessment framework to systematically evaluate the knowledge coverage, logical reasoning abilities, and scenario adaptability of LLMs in the context of carbon footprint management. The analysis reveals significant differences in capabilities, identifies current limitations, and outlines potential improvement directions for mainstream large language models (such as ChatGPT-4o, DeepSeek-R1, and others) in addressing the unique challenges of the carbon footprint domain.

EVALUATING THE CAPABILITIES OF LLMS IN CARBON FOOTPRINT MANAGEMENT

This study employs the MetaEvaluation scoring model to construct a multi-dimensional evaluation framework for assessing large LLMs) in carbon footprint question-answering tasks. The test set comprises three levels of questions: basic questions, which examine knowledge of terminology and conceptual understanding; intermediate questions, which test logical reasoning and computational ability; and advanced integrated questions, which evaluate multi-factor coupling and scenario adaptability. A total of ten representative questions were developed, and responses were evaluated across five dimensions: accuracy (A1), completeness (A2), logical coherence (A3), practicality (A4), and credibility (A5). To reduce evaluator subjectivity, Principal Component Analysis (PCA) was used to determine the dimensional weights. Five mainstream LLMs, including ChatGPT-4o, ChatGPT-o1, DeepSeek-V3, DeepSeek-R1, and Doubao AI were tested under identical prompting conditions. Their results were processed using the TOPSIS method to calculate each model's relative closeness to the ideal solution.

Figure 1 presents the evaluation results of these models in the context of carbon footprint management based on the MetaEvaluation framework. Across ten representative carbon-footprint tasks, DeepSeek-R1 achieved the highest total score (90/100), followed by ChatGPT-o1 (85), DeepSeek-V3 (82), Doubao AI (78), and ChatGPT-4o (71). Overall, the models display evident shortcomings in knowledge coverage, reasoning ability, scenario adaptation, dynamic modeling, and policy sensitivity.

First, LLMs show an incomplete understanding of authoritative standards such as the GHG Protocol and ISO 14064. Their training data often omit recent updates and detailed industry case studies, resulting in ambiguous terminology and limited representativeness. Second, logical reasoning and multivariate analysis remain critical weaknesses. For instance, when addressing complex causal relationships or supply chain interconnection analyses, models frequently make classification errors or exhibit logical biases.

Moreover, scenario adaptation and tool development capabilities are limited. In practical tasks, such as calculating transportation fuel consumption or resolving challenges related to regional regulations, LLMs often struggle to provide actionable solutions. This limitation arises from insufficient exposure to real-world business scenarios during training, such as dynamic

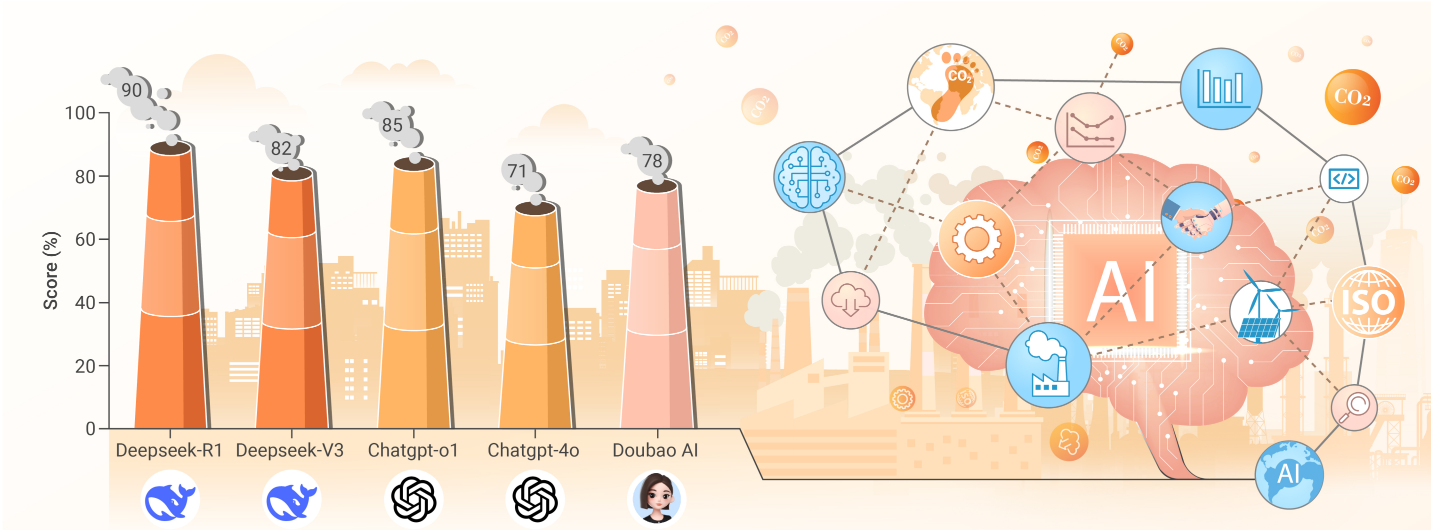

Figure 1. Assessment of mainstream LLMs capabilities in the field of carbon footprints.

cost models or correction factor tables. Additionally, limited policy sensitivity and regional flexibility present major obstacles. For example, LLMs fail to track and analyze international policy changes (e.g., CBAM offset regulations) in real time or address the nuanced challenges of policy implementation in underdeveloped regions.

Beyond these general shortcomings, specific performance differences are evident among models. Among the five weighted evaluation dimensions, accuracy (36%), logical coherence (32%), completeness (23%), practicality (7%), and credibility (3%), ChatGPT excels at interpreting terminology and addressing low to medium-complexity queries. However, it underperforms in advanced logical reasoning and dynamic data processing. In contrast, DeepSeek demonstrates strong logical analysis capabilities for medium to high-complexity problems. Nevertheless, it lags in tool development and policy adaptability, particularly in addressing dynamic regulatory frameworks and real-world applications. Doubao AI exhibits intermediate overall performance, its strengths lie in generating structured answers and maintaining semantic coherence, it shows clear deficiencies in domain knowledge coverage, reasoning depth, and dynamic scenario modeling.

FUTURE PERSPECTIVE FOR LLMS IN CARBON FOOTPRINT MANAGEMENT

To enhance the capabilities of LLMs in carbon footprint management, future improvements should focus on five key areas. First, optimizing and integrating knowledge bases is essential. LLMs need to incorporate the latest updates from international standards, such as the GHG Protocol and ISO 14064, along with industry-specific case studies. Building a comprehensive terminology and scenario mapping library will improve the coverage of specialized sectors and life cycle stages, ensuring more accurate and relevant outputs.

Second, enhancing logical reasoning abilities is critical. Structured parsing modules for standard documents should be developed, focusing on refining elements like "uncertainty analysis." Probabilistic models can help address logical inconsistencies in supply chain variables, enabling LLMs to handle complex causal relationships and multivariate challenges with greater precision.

Third, LLMs must optimize their tool and strategy generation capabilities. By integrating dynamic risk assessment tools, such as carbon tax delay probability models, and providing quantitative templates like correction factor tables and standardized process flowcharts, LLMs can significantly improve the efficiency of carbon accounting and emission reduction planning for enterprises.

Fourth, improving policy adaptation and regional identification is crucial. Real-time access to policy databases and the development of regional feature identification models will allow LLMs to dynamically update international regulations, such as CBAM (Carbon Border Adjustment Mechanism) and CCA (Climate Change Agreements). These enhancements will enable the generation of tailored policy recommendations that address specific regional needs and regulatory differences.

Finally, advancing dynamic modeling and multivariate analysis capabilities is necessary. Multivariate linkage modules should be developed to simulate the combined effects of factors like carbon prices, exchange rates, and policy changes. Improved parameter quantification training, such as modeling CCUS (Carbon Capture, Utilization, and Storage) capture rate ranges, will further support scenario analysis and optimize decision-making in complex conditions.

By addressing these areas, LLMs will evolve to demonstrate greater intelligence and efficiency in carbon accounting, emission reduction planning, and policy analysis. These advancements will play a vital role in supporting global efforts to achieve carbon neutrality and tackle climate change effectively.

CONCLUSION

This paper systematically evaluates the performance of various large language models in professional question-answering within the field of carbon footprint management using the MetaEvaluation scoring model. The analysis highlights differences in model capabilities, identifies key shortcomings, and proposes targeted improvement directions. As large language model technology continues to evolve rapidly, it is essential to regularly adjust evaluation dimensions and weights to align with technological advancements and domain-specific requirements, ensuring the assessment remains both adaptable and accurate.

The findings of this study provide valuable insights for the practical application of large language models in carbon footprint management and offer ideas for further research and optimization. With continuous improvements in knowledge integration, enhanced logical reasoning and dynamic analysis capabilities, and greater adaptability to policy requirements, large language models hold immense potential to become intelligent support tools. These advancements will significantly contribute to achieving global carbon neutrality goals and addressing the pressing challenges of climate change.

REFERENCES

1. Cannon A.J. (2025). Twelve months at 1.5 °C signals earlier than expected breach of Paris Agreement threshold. *Nat. Clim. Chang.* **15**: 266–269. DOI:10.1038/s41558-025-02247-8
2. Bertolini M., Duttilo P. and Lisi F. (2025). Accounting carbon emissions from electricity generation: A review and comparison of emission factor-based methods. *Appl. Energy* **392**:125992. DOI:10.1016/j.apenergy.2025.125992
3. Zhang X., Guo X., Zhao J., et al. (2025). Intelligent application of large language model to life cycle assessment methodology. *J. Clean. Prod.* **529**:146776. DOI:10.1016/j.jclepro.2025.146776
4. Tabebordbar A., Beheshti A., Benatallah B., et al. (2020). Feature-Based and Adaptive Rule Adaptation in Dynamic Environments. *Data Sci. Eng.* **5**:207–223. DOI:10.1007/

s41019-020-00130-4
5. Yu F., Yuan Q., Sheng X., et al. (2024). Understanding carbon footprint: An evaluation criterion for achieving sustainable development. *Chinese J. Popul. Resour. Environ.* **22**:367–375. DOI:10.1016/j.cjpre.2024.11.001
6. Moravec V., Gavurova B. and Kovac V. (2025). Environmental footprint of GenAI – Changing technological future or planet climate. *J. Innov. Knowl.* **10**:100691. DOI:10.1016/j.jik.2025.100691
7. Han T., Cong R.G., Yu B., et al. (2024). Integrating local knowledge with ChatGPT-like large-scale language models for enhanced societal comprehension of carbon neutrality. *Energy AI* **18**:100440. DOI:10.1016/j.egyai.2024.100440

FUNDING AND ACKNOWLEDGMENTS

This paper was supported by China University of Petroleum (Beijing) project no. 2462023YJRC009.

DECLARATION OF INTERESTS

The authors declare no competing interests.