Advances in Geo-Energy Research

Vol. 20, No. 3, p. 201-204, 2026

Perspective

Advances in physics-constrained and data-driven dual paradigm for artificial intelligence in oil and gas

Gang Hui[1], Muming Wang[2,3]*, Haibo Cheng[4]*

[1]State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing 102249, P. R. China

[2]East China Petroleum Bureau of China Petroleum & Chemical Corporation, Nanjing 210019, P. R. China

[3]Department of Chemical and Petroleum Engineering, University of Calgary, Calgary, AB, T2N 1N4, Canada

[4]State Key Laboratory of Robotics and Intelligent Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, P. R. China

Keywords:
Physics-constrained
data-driven
knowledge-based
geological modelling

Cited as:
Hui, G., Wang, M., Cheng, H. Advances in physics-constrained and data-driven dual paradigm for artificial intelligence in oil and gas. Advances in Geo-Energy Research, 2026, 20(3): 201-204.
https://doi.org/10.46690/ager.2026.06.01

Abstract:
Integrating physical mechanisms with data-driven methods overcomes the limitations of purely data-driven artificial intelligence and purely mechanism-based models. Purely data-driven approaches suffer from poor interpretability and weak generalization under sparse data, while purely physics-based models are computationally expensive and struggle with complex nonlinearities. This work highlights advances in the physics-constrained, data-driven dual paradigm across petroleum engineering: mechanism–artificial intelligence fusion via Bayesian networks provides traceable hydrocarbon spatial distribution predictions; knowledge–data-driven modelling ensures geological realism; and collaborative physics–data fault diagnosis enhances well monitoring under noise. These advances demonstrate that deep fusion of domain knowledge, physical laws, and multi-source data is essential for creating interpretable, reliable, and efficient intelligent systems for complex subsurface resource development.

1. Introduction

Against the current backdrop of global energy transition and intelligent development, artificial intelligence (AI) technology is integrating deeply into various aspects of oil and gas exploration and development at an unprecedented pace (Cao et al., 2025; Sinha and Dindoruk, 2025). From seismic interpretation and reservoir characterization to drilling optimization, production forecasting, and well integrity monitoring, AI-driven approaches are increasingly employed to address complex subsurface challenges and optimize hydrocarbon recovery (Wang et al., 2025). These challenges include strong geological heterogeneity, multiphase flow nonlinearities, and significant uncertainty in subsurface properties. However, despite their successes, purely data-driven methods suffer from poor interpretability, lack of physical consistency, and weak generalization under sparse data conditions (Hui et al., 2021). Because these methods rely heavily on large, high-quality labeled datasets, they often fail to extrapolate beyond the training regime, producing results that may violate fundamental physical laws. Conversely, conventional mechanism-based models, while physically grounded and interpretable, are computationally expensive and struggle to capture the complex, multi-scale nonlinear patterns inherent in subsurface systems (Zhuang et al., 2025). The governing partial differential equations require high-resolution spatial and temporal discretization, leading to prohibitive computational costs for field-scale applications, and simplifying assumptions inevitably introduce model bias. Recognizing the complementary nature of these two paradigms, the integration

*Corresponding author.
E-mail address: hui.gang@cup.edu.cn (G. Hui); muming.wang@ucalgary.ca (M. Wang); chenghaibo@sia.cn (H. Cheng).

Yandy Scientific Press

Received April 19, 2026; revised April 30, 2026; accepted May 7, 2026; available online May 9, 2026.

of physical constraints with data-driven methodologies has emerged as a promising solution (Karagiorgi et al., 2022; Nagao et al., 2024; Hussain et al., 2025). This integration aims to embed domain knowledge, conservation laws, and governing equations into machine learning frameworks—for example, through physics-informed neural networks, Bayesian inference with physics-based priors, or hybrid modelling that couples reduced-order physics with data-driven corrections. Such synergy preserves the interpretability and robustness of physics-based models while leveraging the flexibility and pattern-recognition capabilities of data-driven approaches. This work highlights key advances in physics-constrained and data-driven AI applications within petroleum engineering, covering three representative areas: mechanism-AI fusion for resource spatial distribution, which ensures traceable and geologically consistent predictions; knowledge- and data-driven geological modelling, which enforces sedimentary realism and physical law adherence; and collaborative physics-data driven fault diagnosis, which enhances well monitoring reliability under noisy field conditions.

2. Applications

2.1 Mechanism-AI fusion for resource spatial distribution prediction

Predicting the spatial distribution of oil and gas resources remains a challenging task due to the complex, nonlinear relationships among multiple accumulation-controlling factors (Ren et al., 2026). Pure mechanism-based modelling relies heavily on expert judgment and suffers from strong subjectivity, while purely data-driven approaches lack geological constraints and often produce results that are difficult to interpret. The dual-drive approach integrating hydrocarbon accumulation mechanisms and artificial intelligence offers an efficient solution for this problem.

Taking hydrocarbon accumulation mechanisms as physical constraints, this method quantitatively characterizes key accumulation-controlling factors including source rocks, reservoirs, cap rocks and traps. By fusing multi-source data such as exploration wells and geological grids, predictions are implemented using Bayesian network classifiers and optimized models (Fan et al., 2026). The Bayesian network visually represents the interdependencies among geological elements through directed acyclic graphs and quantifies the contribution of each element to hydrocarbon accumulation via conditional probabilities. This framework possesses inherent physical interpretability, ensuring that results are traceable and verifiable. Applications in the Bohai Bay Basin and Junggar Basin have effectively improved prediction accuracy, showing remarkably superior performance to traditional methods. This approach allows precise delineation of favorable zones for remaining hydrocarbons. More broadly, multimodal architectures that integrate 3D geological models with tabular and time-series data have also demonstrated superior accuracy and stability in predicting production from unconventional reservoirs (Wang et al., 2025), highlighting the value of fusing diverse data modalities in subsurface applications.

2.2 Knowledge- and data-driven geological modelling

Purely data-driven geological modelling methods often suffer from poor generalization when training data are sparse, and they lack the ability to enforce geological realism (Bhavsar et al., 2024). Without proper constraints, these models may generate geologically implausible structures that deviate from established sedimentary principles (Bao et al., 2025). Conversely, purely knowledge-driven approaches, while geologically sound, are often too rigid to capture local variations observed in actual field data. Knowledge- and data-driven geological modelling methods address these limitations by integrating prior geological patterns and physical equation constraints with data mining capabilities.

This approach establishes a surrogate model for geological modelling that features conditional data compliance, geological feature similarity and adherence to physical laws. During surrogate model training, physics-informed neural networks (PINNs) are introduced to enforce compliance with physical laws by embedding governing equations into the loss function. Similar PINN frameworks have also been successfully applied to reservoir simulation, where incorporating governing seepage equations into the loss function ensures physical consistency and improves generalization beyond training data (Sinha and Dindoruk, 2025). Additionally, geological images serve as training data to reinforce prior geological constraints, and hard-constraint modules for conditional data embedding are adopted. These measures ensure that the generated geological model possesses high geological realism and physical interpretability. The synergistic integration of knowledge and data has become a fundamental paradigm in geological modelling, although practical application still requires consideration of multi-physics coupling and real-time dynamic updating mechanisms.

2.3 Collaborative physics-data driven fault diagnosis for production wells

Fault diagnosis for production wells, particularly for beam pumping units, is critical for maintaining field production efficiency. Traditional data-driven methods have two major weaknesses: they are susceptible to signal distortion under strong noise interference, and they rely heavily on handcrafted feature extraction (He et al., 2024). In many oilfield environments, the collected dynamometer cards are often contaminated by various noise sources, making accurate diagnosis challenging. Production well fault diagnosis empowered by the co-driving of physical mechanisms and data has demonstrated significant advantages in overcoming these limitations.

By integrating the structural dynamic mechanism of beam pumping units with edge-intelligent deep learning algorithms, a synergistic framework combining physics-constrained data purification with end-to-end fault classification has been established. This collaborative physics-data driven paradigm substantially improves the accuracy of fault diagnosis while simultaneously enhancing the physical interpretability of the model. Moreover, it satisfies the real-time requirements of edge computing in the Industrial Internet of Things, making it suitable for field deployment. Overall, future intelligent monitoring of oil and gas equipment should further strengthen the deep coupling of physical mechanisms with multi-source sensing data to promote the development of highly trustworthy edge-intelligence systems for complex operating conditions.

3. Challenges

While current challenges are often broadly acknowledged, a deeper analysis of their root causes is essential. The computational difficulty comes from two sources. One is high dimensionality. But the more fundamental cause is the stiff, nonlinear nature of the governing partial differential equations themselves. These equations require highly stable numerical schemes, which are often at odds with the gradient-based optimization used in machine learning. Similarly, the challenge in balancing data-fit and physics-based losses arises from a profound methodological gap: the absence of a principled framework to reconcile statistical deviations with mechanistic constraints, especially under sparse or noisy data regimes. This frequently leads to adversarial learning dynamics where improving physical consistency may compromise predictive accuracy, and vice versa. In highly heterogeneous reservoirs, the intrinsic multiscale nature of flow and transport phenomena further complicates the faithful encoding of physical laws into scalable loss functions.

Looking ahead, multimodal machine learning offers promising pathways for modelling complex subsurface systems through the fusion of diverse data types. A critical challenge remains. Domain knowledge must be incorporated into model architecture and validation processes. This incorporation needs to be effective and systematic. Only then can physical plausibility and practical relevance be ensured across varied field applications. More notably, large multimodal models based on transformer architectures are breaking down scale barriers between core, logging, and seismic data through cross-modal attention mechanisms. Microscopic core features can directly constrain macroscopic physical inversion processes, enabling pore-permeability parameters obtained at the laboratory scale to participate directly in reservoir numerical simulation and hydraulic fracturing optimization design. However, the computational cost and data requirements of these large models present new obstacles for practical field deployment.

4. Conclusions

The integration of physics constraints with data-driven AI technologies is reshaping oil and gas exploration and development by addressing key limitations of purely data-driven or purely mechanism-based approaches. The advances presented herein cover three key areas: mechanism-AI fusion for resource spatial distribution, knowledge- and data-driven geological modelling, and collaborative physics-data driven fault diagnosis. Taken together, these advances demonstrate the dual paradigm's capacity to handle complex, multi-source data while maintaining physical consistency and interpretability. These developments highlight that the future of AI in petroleum engineering lies in the deep fusion of domain expertise, physical laws, and diverse data sources. Such integration ultimately supports more efficient, reliable, and sustainable oil and gas resource development within the context of the global energy transition.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 62203432), the Research Program of Liaoning Liaohe Laboratory (No. LLL25ZZ-05-01) and the National Major Science and Technology Program (No. 2025ZD1401405).

Conflicts of interest

The authors declare no competing interests.

References

Bao, P., Hui, G., Hu, Y., et al. Comprehensive characterization of hydraulic fracture propagations and prevention of pre-existing fault failure in Duvernay shale reservoirs. Engineering Failure Analysis, 2025, 173: 109461.

Bhavsar, F., Desassis, N., Ors, F., et al. A stable deep adversarial learning approach for geological facies generation. Computers & Geosciences, 2024, 190: 105638.

Cao, L., Jiang, F., Chen, Z., et al. Data-driven interpretable machine learning for prediction of porosity and permeability of tight sandstone reservoir. Advances in Geo-Energy Research, 2025, 16(1): 21-35.

Fan, X., Zhao, F., Yu, J., et al. DWOA-BNC: Discrete whale optimization algorithm for Bayesian network classifier learning and its application. Pattern Recognition, 2026, 173: 112661.

He, Y., Cheng, H., Zeng, P., et al. Working condition recognition of sucker rod pumping system based on 4-segment time-frequency signature matrix and deep learning. Petroleum Science, 2024, 21(1): 641-653.

Hui, G., Chen, S., He, Y., et al. Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors. Journal of Natural Gas Science and Engineering, 2021, 94: 104045.

Hussain, A., Pan, P. Z., Hussain, J., et al. Data-driven machine learning models for predicting deliverability of underground natural gas storage in aquifer and depleted reservoirs. Energy, 2025, 319: 134974.

Karagiorgi, G., Kasieczka, G., Kravitz, S., et al. Machine learning in the search for new fundamental physics. Nature Reviews Physics, 2022, 4(6): 399-412.

Nagao, M., Datta-Gupta, A., Onishi, T., et al. Physics informed machine learning for reservoir connectivity identification and robust production forecasting. SPE Journal, 2024, 29(9): 4527–4541.

Ren, H., Fan, X., Liang, K., et al. Prediction method for the spatial distribution of oil and gas resources based on a

Bayesian network classifier: A case study of the Shahejie formation in southeastern Jizhong Depression, Bohai Bay Basin, China. Mathematical Geosciences, 2026, 58(2): 381-402.

Sinha, U., Dindoruk, B. Review of physics-informed machine learning (PIML) methods applications in subsurface engineering. Geoenergy Science and Engineering, 2025, 250: 213713.

Wang, M., Wang, H., Hui, G., et al. A hybrid tabular-spatial-temporal model with 3D Geomodel for production prediction in shale gas formations. SPE Journal, 2025, 30(6): 3281-3293.

Zhuang, X., Liu, Y., Hu, Y., et al. Prediction of rock fracture pressure in hydraulic fracturing with interpretable machine learning and mechanical specific energy theory. Rock Mechanics Bulletin, 2025, 4(2): 100173.