BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

ARTIFICIAL INTELLIGENCE AND BIG DATA IN THE OIL AND GAS INDUSTRY: SEISMIC INTERPRETATION, DRILLING-RISK PREDICTION, DIGITAL TWINS, AND PREDICTIVE MAINTENANCE

Ro'zimova Komila Komil qizi
Student, Mathematics Program, Faculty of Mathematics,
National University of Uzbekistan named after Mirzo Ulugbek
E-mail: ruzimovak@gmail.com
doi:https://doi.org/10.5281/zenodo.24275975

Abstract. This article critically examines the methodological integration of artificial intelligence and Big Data within oil-and-gas operations. It connects deep-learning-based seismic interpretation, early identification of drilling hazards, digital-twin modeling, and prediction of equipment degradation within a unified engineering decision framework. Published studies by Uzbekistan-based scholars are compared with international research, with particular attention to data quality, physically admissible modeling, external validation, uncertainty, and industrial safety. The article proposes a layered conceptual architecture and task-specific evaluation criteria. It is an integrative literature study; the results describe a reasoned analytical synthesis rather than newly collected field measurements or experimentally demonstrated performance improvements.

Keywords: artificial intelligence, Big Data, oil and gas, seismic interpretation, neural networks, drilling risk, digital twins, predictive maintenance.

INTRODUCTION

The convergence of artificial intelligence (AI) and Big Data has altered the epistemological and technological foundations of the contemporary oil and gas industry. Exploration, drilling, production, transportation, and maintenance generate heterogeneous information whose value depends less on its raw volume than on the capacity to transform it into validated engineering knowledge. Three-dimensional seismic surveys, well logs, drilling-control systems, distributed pressure and temperature sensors, geological simulations, and maintenance records differ in sampling frequency, spatial resolution, uncertainty, and physical meaning. Their joint interpretation consequently demands more than the mechanical application of algorithms. A scientifically defensible analytical system must integrate statistical learning with petroleum geology, reservoir physics, numerical mathematics, and operational safety. In this context AI is not merely a tool for automating routine calculations; it is an instrument for identifying nonlinear relations, representing uncertainty, and supporting decisions across the lifecycle of hydrocarbon assets.

The Big Data paradigm is commonly associated with volume, velocity, variety, veracity, and value. In petroleum operations each dimension has a distinctive consequence. Large seismic volumes require scalable storage and computation; high-frequency drilling signals call for near-real-time analysis; laboratory measurements and textual incident reports introduce structural heterogeneity; missing or miscalibrated sensors undermine veracity; and economic value appears only when analyses improve identifiable decisions. For example, a model that classifies well conditions with impressive accuracy but cannot trigger a safe intervention has limited practical utility. Data governance, physical consistency, temporal traceability, and domain-specific validation must therefore precede claims concerning intelligent oilfield management. The task is particularly consequential because false negatives in drilling-hazard detection may have direct implications for personnel safety, equipment integrity, and environmental protection.

Seismic interpretation illustrates the relationship between mathematical abstraction and geological expertise. Reflection signals contain information about interfaces, stratigraphic discontinuities, faults, and structural traps, but seismic amplitudes do not uniquely determine lithology or hydrocarbon saturation. Noise, migration artifacts, acquisition footprints, and variations in wave propagation may resemble geological features. Conventional interpretation relies on specialist inspection and engineered attributes such as coherence or semblance; contemporary convolutional neural networks can learn multi-scale patterns and delineate fault likelihoods throughout a seismic volume. The FaultSeg3D work of Wu and colleagues demonstrated the methodological significance of training a segmentation network on synthetic three-dimensional examples [3]. Nevertheless, such predictions remain probabilistic interpretations. They must be compared with well control, structural plausibility, and uncertainty estimates rather than presented as direct measurements of recoverable reserves.

Drilling operations generate a second, fundamentally temporal, type of Big Data. Weight on bit, rotary speed, torque, standpipe pressure, mud flow, vibration, temperature, and rate of penetration evolve under changing lithology and operating regimes. Aberrations can indicate stuck-pipe tendencies, lost circulation, pump degradation, or other hazardous events, yet normal transitions may produce similar signal patterns. Supervised classification, anomaly detection, temporal neural networks, and survival models can help rank emerging risks if their training data describe genuine incidents and valid operating contexts. Wu and coauthors' study of lost-circulation risk emphasizes the need to connect statistical recognition to drilling-specific hazard severity [4]. A useful warning system must be calibrated, interpretable to engineers, sensitive to rare high-consequence events, and evaluated with historical sequences that preserve the temporal order of operations.

LITERATURE REVIEW

The theoretical literature can be organized around two related problems: learning the state of subsurface or mechanical systems from incomplete observations and converting statistical signals into defensible decisions. Among Uzbek researchers, S. S. Muxsinov’s article on artificial intelligence in forecasting gas-field development indicators is especially pertinent [1]. Published in 2025, it examines a Long Short-Term Memory (LSTM) recurrent neural-network approach using geophysical and production-related information obtained from historical hydrodynamic simulation. The contribution lies in selecting a sequence-sensitive architecture for indicators that evolve through time. Unlike a static regression equation, an LSTM model can represent dependencies between successive operating states. Its methodological implication is that a well forecast should preserve history, not merely aggregate mean conditions. At the same time, model training on simulation-generated trajectories does not automatically establish performance under sensor failure, unmodeled interventions, or a radically different reservoir. The distinction between an internally accurate forecast and a transferable operational predictor remains decisive.

A second Uzbek contribution is the 2026 study by Mirzohid Murodovich Abdukarimov and Behro’z Sanjar o‘g‘li Dilmurodov, which addresses AI-supported management of oil and gas fields [2]. The authors describe SCADA streams from 24 wells in three fields in the Bukhara–Khiva and Ustyurt regions and employ data preprocessing and discriminant analysis to distinguish operating conditions. The published abstract reports classification accuracy ranging from 93.1% to 96.2% across three classes; those figures belong to their own study and are not outcomes generated for the present paper. Their work illustrates the operational usefulness of comparatively transparent statistical classification, especially when continuous supervision requires rapid differentiation among well states. In methodological terms, their contribution complements Muxsinov’s

forecasting study: one emphasizes temporal prediction, whereas the other emphasizes state recognition and decision-oriented categorization.

The two national studies should not be read as competing demonstrations of algorithmic superiority because their prediction targets, data-generating mechanisms, and evaluation arrangements differ. Their shared significance lies in establishing a bridge between industrial data acquisition and mathematical inference in Uzbekistan. Together they raise the question of what happens after a forecast or a classification is produced: can the result be traced to an intervention, assessed for uncertainty, and transferred to another field? This article treats that question as the missing systems-level connection among local applications. It is particularly relevant for machine-learning systems trained on geographically concentrated datasets, where geological differences and operational policy changes can create silent performance degradation.

International evidence broadens the methodological scope. Xinming Wu and collaborators formulated three-dimensional seismic fault extraction as a supervised image-segmentation task [3], while Shengnan Wu and colleagues analyzed risk identification for lost circulation during drilling [4]. Raissi, Perdikaris, and Karniadakis proposed the widely used physics-informed neural-network framework [5]. Mei Wang and coauthors reported an ensemble-based approach to fault prediction from oilfield equipment data [6]; Jiangang Wang's team developed a digital-twin framework for a rig lifting system [7]; and Zhang and collaborators reviewed decision-oriented wellbore twins [8]. Recent work on drilling mud-pump maintenance [9] and the hybrid role of geological interpretation methods [10] further indicates that an effective digital strategy must accommodate distinct model classes instead of replacing engineering expertise with a single universal network.

METHODOLOGY

This study employed an integrative literature-review method combined with comparative analysis, systems modeling, mathematical formalization, and scenario-based evaluation. The evidence base was restricted to ten identifiable academic publications with verifiable bibliographic metadata, including two contributions by Uzbekistan-based researchers and international articles directly related to seismic interpretation, drilling hazards, digital twins, physics-informed computation, and equipment maintenance [1–10]. Sources were examined according to their research objects, input-data structure, computational approach, validation logic, and claimed operational implications. The analysis differentiated original empirical findings, engineering prototypes, methodological reviews, and conceptual proposals. No proprietary oilfield database was accessed and no new experiments or production measurements were conducted for this article. Therefore the results presented below are a reproducible analytical synthesis and a proposed implementation framework, not newly measured predictive performance.

RESULTS

The principal result of the analytical synthesis is a five-layer functional architecture for AI and Big Data in the petroleum value chain. The foundational layer concerns acquisition and governance: seismic surveys, downhole logs, SCADA streams, maintenance histories, and numerical simulations must be stored with provenance, timestamps, spatial coordinates, units, and quality indicators. The second layer performs data reconciliation and feature construction, while the third contains task-specific learning models. The fourth integrates these models with reservoir and equipment dynamics in a continuously revised digital representation. The fifth translates verified outputs into ranked operational recommendations, subject to explicit human review. This layered interpretation is a conceptual outcome of the present study; it is not evidence that a new integrated platform has already been deployed in a field.

For seismic analysis, the review identifies three necessary forms of validation: pixel- or voxel-level agreement with expert-labeled faults, consistency of inferred structures across three-dimensional space, and plausibility relative to available well and geological evidence. Wu and colleagues' FaultSeg3D provides a documented example of segmentation-based deep learning using synthetic training volumes [3]. That evidence supports the feasibility of automated structural delineation but does not imply direct determination of hydrocarbon reserves from seismic amplitude alone. A proposed oilfield workflow would therefore assign probability maps to interpreters, preserve alternative structural scenarios, and carry uncertainty into subsequent reservoir simulation. Classification confidence and geologic validity are different properties and should be reported separately.

For drilling-risk prediction, the synthesis establishes that useful forecasting requires calibrated probabilities and adequate early-warning time, not simply high overall classification accuracy. Lost-circulation hazards provide a published example of event-oriented machine-learning analysis [4]. The proposed risk engine combines drilling-fluid pressure, flow imbalance, torque, vibration, and operational context within a time-window model; its outputs would include incident likelihood, contributing features, and an advisory severity category. Because serious events are infrequent, evaluation should emphasize missed events, alarm burden, and cross-well validation. A model with excellent retrospective aggregate accuracy could still be unsafe if it fails specifically during unusual formations or equipment transitions. No numerical performance is claimed for the proposed engine because its operational dataset has not been assembled here.

For production forecasting and digital twins, two complementary pathways emerged. Muxsinov's published LSTM application supports sequence-based modeling of field development variables [1], whereas Abdukarimov and Dilmurodov demonstrate classification of well operating conditions using SCADA-derived features [2]. The latter study reports accuracies from 93.1% to 96.2%, which are cited solely as its reported findings. In the proposed architecture, temporal forecasts and current-state labels become separate inputs to a digital twin; each must remain distinguishable from measured pressures, rates, and other physical quantities. Published drilling-rig twin work [7] and the wellbore digital-twin review [8] support integrating mechanistic models and updating the virtual state as observations arrive, but they also indicate that data-fusion quality and model synchronization are continuing technical challenges.

DISCUSSION

The central methodological debate can be articulated through two internationally influential perspectives: the data-driven seismic interpretation advanced by Xinming Wu and the physics-informed learning associated with George Em Karniadakis. The present discussion constructs a comparison between their published approaches; it does not claim that these researchers engaged in a direct personal controversy. Wu's work on FaultSeg3D treats a geological interpretation problem as three-dimensional semantic segmentation and demonstrates how synthetic labeled volumes can support a convolutional network when large collections of annotated field images are unavailable [3]. This orientation privileges representational learning and the capacity of networks to extract complex spatial features. Its attraction is considerable: once a useful model has been trained, structurally coherent fault predictions may be generated at scale, reducing repetitive interpretation effort. The difficulty concerns the relationship between the synthetic training distribution and the heterogeneous geology of real basins.

Karniadakis and coauthors, in contrast, emphasize embedding governing equations within the learning process [5]. Their argument is not that physical modeling can dispense with observations, but that unconstrained fit may be insufficient when samples are sparse, noisy, or extrapolative.

Applied to a reservoir or wellbore, this logic encourages consistency with conservation laws, flow behavior, boundary conditions, and measurable operating limits. A physics-informed loss can disfavor predictions that reproduce past data while violating known dynamics. Yet physical models are also imperfect: multiphase flow, fractures, heterogeneity, and uncertain rock parameters complicate the governing description. A weak or incorrect physical specification may impose bias, producing apparently disciplined but misleading results. Physics-based regularization is therefore a conditional resource, not a guarantee of correctness.

CONCLUSION

The present review concludes that artificial intelligence and Big Data are most valuable in the oil-and-gas industry when designed as a coordinated engineering decision system rather than as a collection of unrelated prediction tools. Seismic deep learning can assist fault delineation; temporal models can recognize changes in drilling or production states; digital twins can connect observations to physics-based virtual representations; and predictive maintenance can guide repair timing through condition and risk estimates. Their integration depends on well-documented datasets, consistent units and timestamps, independent field validation, explicit uncertainty quantification, and operator-accessible explanations. The approaches of Uzbek researchers S. S. Muxsinov and M. M. Abdukarimov, together with B. S. Dilmurodov, provide locally relevant starting points for further comparative work.

The article proposes a phased implementation sequence: first, develop a trustworthy data foundation; second, benchmark task-specific algorithms against transparent baselines; third, test hybrid physical–statistical models on held-out wells and time periods; fourth, synchronize the validated models within bounded digital-twin applications; and fifth, assess maintenance and safety decisions through operational outcomes rather than accuracy alone. Priority areas for subsequent empirical work include seismic fault interpretation under geological domain shift, early warnings for rare drilling events, calibration of remaining-useful-life forecasts, and evaluation of digital-twin uncertainty. Because this paper is an integrative literature study, it claims no new field measurements, no independently achieved productivity gains, and no experimentally verified cost reductions. Such effects require prospective trials with clear comparators and auditable evidence.

REFERENCES

1. Muxsinov S. S. Применение искусственного интеллекта в прогнозировании показателей разработки газового месторождения // Fan va texnologiyalar taraqqiyoti. — 2025. — № 4. — С. 132–138. — URL: https://journal.bstu.uz/public/article/227/ (accessed: 16.11.2026).
2. Abdukarimov M. M., Dilmurodov B. S. Sun‘iy intellekt yordamida neft va gaz konlarini boshqarish // Бюллетень студентов Нового Узбекистана. — 2026. — Vol. 4, № 6. — P. 205–212. — DOI: 10.70413/yota-v04-i06-s54144.
3. Wu X., Liang L., Shi Y., Fomel S. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation // Geophysics. — 2019. — Vol. 84, № 3. — P. IM35–IM45. — DOI: 10.1190/geo2018-0646.1.
4. Wu S., Hu Y., Zhang L. et al. Intelligent risk identification for drilling lost circulation incidents using data-driven machine learning // Reliability Engineering & System Safety. — 2024. — Vol. 252. — Art. 110407. — DOI: 10.1016/j.ress.2024.110407.
5. Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // Journal of Computational Physics. — 2019. — Vol. 378. — P. 686–707. — DOI: 10.1016/j.jcp.2018.10.045.

6. Wang M., Su X., Song H. et al. Enhancing predictive maintenance strategies for oil and gas equipment through ensemble learning modeling // Journal of Petroleum Exploration and Production Technology. — 2025. — Vol. 15. — Art. 46. — DOI: 10.1007/s13202-025-01931-x.
7. Wang J., Shi L., Ding F. et al. A digital twin modeling and application for gear rack drilling rigs lifting system // Scientific Reports. — 2024. — Vol. 14. — Art. 23711. — DOI: 10.1038/s41598-024-73954-z.
8. Zhang F., Zhang C., Wang X. et al. Digital twin technology for wellbore and intelligent decision-making in drilling operations: Current status and future prospects // Petroleum Science Bulletin. — 2026. — Vol. 11, № 1. — P. 191–208. — DOI: 10.3969/j.issn.2096-1693.2026.01.002.
9. Wu X., Chang J., Long C., Ghorbani H. Machine learning insight in quality control and maintenance management of drilling mud pumps // Journal of Big Data. — 2026. — Vol. 13. — Art. 132. — DOI: 10.1186/s40537-026-01475-9.
10. Wu X., Fomel S. Automatic fault interpretation with optimal surface voting // Geophysics. — 2018. — Vol. 83, № 5. — P. O67–O82. — DOI: 10.1190/geo2018-0115.1.