

Invited review

Large language models and foundation models in oil and gas exploration and development

Zhangxing Chen¹, Yalong Li¹^{*}, Ruichen Ding^{1,2}, Benjieming Liu^{1,3}, Mohammadali Ahmadi⁴, Wei Xiong⁵*

¹Ningbo Key Laboratory of Low-Carbon Hydrogen Energy, Eastern Institute of Technology, Ningbo, Ningbo 315200, P. R. China

²Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, P. R. China

³State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing 102249, P. R. China

⁴Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran 1458889694, Iran

⁵Research Institute of Petroleum Exploration and Development, Beijing 100083, P. R. China

Keywords:

Large language models
foundation models
multimodal learning
AI agents
digital twins
exploration and development

Cited as:

Chen, Z., Li, Y., Ding, R., Liu, B., Ahmadi, M., Xiong, W. Large language models and foundation models in oil and gas exploration and development. *Advances in Geo-Energy Research*, 2026, 22(1): 11-25.
<https://doi.org/10.46690/ager.2026.10.02>

Abstract:

In oil and gas exploration and development, large language models and foundation models offer a new basis for integrating knowledge, data and engineering workflows. However, fragmented applications and unclear distinctions between model capabilities obscure how these technologies can support integrated decision-making. This review synthesizes the development and upstream applications of these models through a layered framework linking multimodal representation, domain knowledge, engineering computation, and operational decision-making. The findings indicate that the principal value of these models lies in connecting established petroleum technologies across disciplinary and organizational boundaries, rather than replacing specialized models or physical understanding. Within this framework, language models support knowledge interaction and workflow coordination, while foundation models extend representation learning across heterogeneous subsurface and operational data. The integration of these models with engineering models provides a pathway from isolated task automation toward coordinated intelligence across the exploration and development life cycle. The central challenge is to translate general model capabilities into reliable, transferable and economically sustainable engineering systems under geological uncertainty and operational constraints. However, addressing this task requires a closer integration of domain knowledge and physical mechanisms, alongside consistent data semantics and workflow-level evaluation. Herein, a staged development pathway is proposed, progressing from scenario-specific assistance to cross-disciplinary collaboration and bounded autonomous operation. This perspective reframes petroleum artificial intelligence as a system-integration problem and provides a conceptual basis for advancing human-machine collaboration in complex energy-resource development.

1. Introduction

The oil and gas industry remains central to global energy security and economic activity. In this context, exploration and development span the fields of geology, geophysics, drilling, reservoir engineering, and production. These activities are not

only capital-intensive but also safety-critical and dependent on heterogeneous subsurface and operational data (Tariq et al., 2021; Dongo and Relvas, 2025). The energy transition and China's carbon-peaking and carbon-neutrality goals further increase the pressure to improve efficiency and reduce envi-

Table 1. Operational taxonomy for the AI model classes discussed in this review.

Technology class	Primary function	Typical petroleum role	Engineering role and evaluation focus
Language-centered LLM	Text understanding, generation, retrieval, and tool calling	Knowledge Q&A, report drafting, code or workflow orchestration	Faithful retrieval, provenance, tool success, abstention, and security
Multimodal foundation model	Representation of images, seismic data, curves, and text	Core-image analysis, seismic interpretation, log representation	Representative benchmarks, cross-survey behavior, uncertainty, and expert correction
Task-specific numerical AI or surrogate	Prediction, classification, approximation, or optimization	Production forecast, history-matching surrogate, equipment diagnosis	Holdout prediction, calibration, physical checks, and operating domain
Digital twin or physics simulator	State estimation and mechanistic scenario evaluation	Reservoir surveillance, drilling simulation, production optimization	Model verification, data assimilation, uncertainty, and version control
LLM-enabled agentic system	Planning and orchestration across governed tools	Multi-step engineering workflow and decision support	Permission controls, execution logs, recovery, operator approval, and safety envelope

ronmental impact (Zhang, 2022; Zhang and Chen, 2022). AI, cloud computing, industrial sensing, and data platforms can support this transformation; however, their value depends on integration with established engineering practice (Mohammadpoor and Torabi, 2020; Sircar et al., 2021; Mirza et al., 2022; Haouel and Nemeslaki, 2024).

Transformer-based large language models (LLMs), including the GPT, BERT, and LLaMA families, have expanded natural language understanding and generation, code synthesis, and tool use. These families build on advances in attention-based architecture, pre-training, instruction tuning, and scale (Devlin et al., 2019; Brown et al., 2020), and their transferability makes them attractive as interfaces to industrial knowledge and software. However, fluent language output is not an evidence of geological validity, numerical accuracy, or operational safety. Petroleum applications therefore require task-specific evaluation, traceable sources, and explicit controls around model actions.

Oil and gas organizations hold large collections of geological reports, seismic and well-log data, as well as drilling records, production histories, technical standards, and expert interpretations (Hassani and Silva, 2018; Koroteev and Tekic, 2021). LLMs can improve access to these textual knowledge and coordinate analytical tools, whereas multimodal foundation models can learn representations from images, curves and seismic volumes (Yin et al., 2024). These functions differ from task-specific predictive models, physics-based simulators and digital twins, which perform numerical estimation or system-state calculations (Di et al., 2021; Li et al., 2024; Liu et al., 2024).

This review examines LLMs and related foundation-model systems in upstream oil and gas exploration and development. It is structured around technology architecture, upstream applications, and deployment requirements, with Table 1 defining the technology classes used throughout the paper to ensure consistent technical attribution and workflow-oriented analysis. The literature included covers studies published from

January 2017 to September 2026 and was drawn from the Web of Science, Scopus, Google Scholar, OnePetro, ScienceDirect, CNKI, and official sources from operators and model providers. Relevant peer-reviewed articles, conference papers, industrial reports, and company announcements were included if they described an upstream task, a model or system configuration, and an evaluation result or deployment context.

2. Development of LLMs: A brief overview

2.1 Evolution of NLP: From statistical methods to neural networks

Natural language processing (NLP) emerged in the 1950s through early work on machine translation and rule-based language systems (Weizenbaum, 1966). In the 1980s, as greater computing capacity and larger text corpora allowed models to learn linguistic patterns from data, the field shifted toward statistical methods (Manning and Schütze, 1999). Statistical NLP supported more scalable and adaptable systems for tasks such as part-of-speech tagging and information extraction.

During the 2000s and 2010s, deep learning and neural networks substantially changed NLP (LeCun et al., 2015). Models increasingly learned distributed representations from data instead of relying on manually specified linguistic rules (Goodfellow et al., 2016). This transition underpinned advances in machine translation, dialogue systems, information retrieval, sentiment analysis, and text generation. The ability of neural networks to learn complex patterns from large datasets subsequently enabled the scale and versatility of modern LLMs (Schmidhuber, 2015).

LLMs are deep neural networks pre-trained on large text corpora, typically using self-supervised objectives. To model relationships among tokens within a sequence, most contemporary LLMs use the Transformer architecture. A single pre-trained model can be adapted to tasks such as question answering, summarization, translation, and text completion, which has expanded its potential use in search, content generation

Table 2. Key milestones in LLMs and foundation models from 2017–2026.

Year	Milestone	Main contribution
2017	Transformer	Establishes an attention-based architecture for scalable sequence modeling (Vaswani et al., 2017).
2018	BERT	Introduces bidirectional masked-language pre-training for transferable language representations (Devlin et al., 2019).
2019	GPT-2	Demonstrates zero-shot language-task transfer through autoregressive pre-training (Radford et al., 2019).
2020	GPT-3	Demonstrates few-shot task adaptation through prompting without updating model weights (Brown et al., 2020).
	RAG	Combines language generation with retrieval from external document memory (Lewis et al., 2020).
2021	CLIP and foundation models	CLIP aligns image and text representations; the foundation-model concept formalizes broad pre-training followed by task adaptation (Radford et al., 2021).
2022	PaLM	Demonstrates large-scale language modeling across multilingual, coding, and reasoning tasks (Chowdhery et al., 2023).
	InstructGPT and ChatGPT	Advanced instruction following through human-feedback training and popularized conversational interaction (Ouyang et al., 2022; OpenAI, 2022).
2023	LLaMA	Released research model weights, expanding opportunities for domain adaptation (Cheng et al., 2024).
	GPT-4 and Gemini	Extended language-centered models to image-text input and broader multimodal processing (Han et al., 2024; Zhang et al., 2024).
	Segment Anything Model (SAM)	Introduced promptable image segmentation using a transferable visual foundation model (Kirillov et al., 2023).
2024	DeepSeek LLM and Gemini 1.5	Advanced dense open language models, long-context multimodal processing, and inference-time reasoning, respectively (OpenAI, 2024).
2025	DeepSeek-R1	Advanced reasoning through reinforcement learning and transferred reasoning capabilities to smaller models (Guo et al., 2025).
	Qwen3, Kimi K2, and GPT-5	Advanced adjustable reasoning, agentic tool use, and routing between fast and deeper-reasoning models, respectively (OpenAI, 2025).
2026	Kimi K2.5, Claude Opus 4.6, and GPT-5.4	Extended vision-agent coordination, long-running workflows, context management, native computer use, and tool search (Anthropic, 2026; OpenAI, 2026a).
	GPT-5.6	Introduces a tiered model family for long-horizon professional workflows, programmatic tool calling, and multi-agent coordination (OpenAI, 2026b).
	GPT-6 Astra	Advanced end-to-end computer use and long-context agentic workflows through asynchronous tool calling and strengthened alignment safeguards (OpenAI, 2026c).

and virtual assistance (Yin et al., 2024).

2.2 Key milestones from 2017 to 2026

In this period, LLMs and foundation models progressed from transferable language representation to multimodal perception, inference-time reasoning, and agentic execution. Table 2 highlights the milestones most relevant to industrial knowledge services, interpretation, and workflow orchestration.

2.3 Core capabilities and industrial value of LLMs

LLMs extend deep-learning approaches to language modeling at very large parameter scales. Pre-training on text and, increasingly, multimodal data can support language under-

standing and generation, knowledge integration, and some forms of reasoning. Relative to task-specific AI models, LLMs offer three frequently cited properties: generalization to new inputs, versatility across tasks, and emergent capabilities that become apparent only beyond certain scales. These properties are not uniform across models or applications and require task-specific evaluation.

General-purpose LLMs can be adapted to petroleum tasks through domain-adaptive pre-training, instruction tuning, parameter-efficient fine-tuning, retrieval-augmented generation (RAG), and integration with knowledge graphs or validated software (Raffel et al., 2020; Liu et al., 2024). RAG is particularly useful when answers must reflect current or organization-specific documents, because retrieved passages can be shown with provenance. Subsequently, fine-tuning may improve ter-

minology and task behavior.

Four model classes recur in the petroleum literature. First, language-centered LLMs support knowledge services and natural-language interaction (Pan et al., 2024; Zhu et al., 2025). Second, multimodal foundation models represent seismic, image, curve, or text data (Zhang et al., 2025a; Gao et al., 2026). Third, task-specific numerical models provide forecasting, classification, surrogate simulation, or control (Cui et al., 2025; Du et al., 2025). Fourth, agentic systems connect an LLM or planner to retrieval, simulators, databases, and optimization routines (Wang et al., 2025; Zhao et al., 2025). The term LLM-enabled system is used here only when a language model provides a substantive semantic or orchestration function. To keep evidence and attribution aligned, each application is interpreted at three levels: demonstrated model use supported by a cited study or deployment; conventional AI, simulation, or optimization that supplies the quantitative calculation; and conceptual LLM-enabled integration proposed as a future or system-level extension.

3. Enabling architectures for petroleum foundation-model systems

3.1 Knowledge services, RAG and organizational memory

One of the most direct routes from LLM capability to petroleum work are knowledge services. Therein, geological reports, technical standards, operating procedures, well histories, and lessons learned can be indexed for natural-language retrieval and question answering. For example, Shengli Oil-field's Sheng Xiaoli system was trained with millions of petroleum-domain records to support information-asset search, drafting, and cross-system data access. Saudi Aramco's 250-billion-parameter MetaBrain system was reported to integrate more than 90 years of company data and analyze drilling plans and geological information when recommending well options (Duan et al., 2024; Sinopec News, 2025c). In both cases, the practical change is a shorter path from user intent to governed organizational memory, allowing younger engineers to retrieve analogue cases and expert knowledge without navigating multiple repositories.

A dependable knowledge service requires document governance, metadata, access control, retrieval evaluation, and expert review. Knowledge graphs can improve entity normalization and relationship search, but they are optional components rather than the defining core of an LLM system. Evaluation should measure retrieval recall, answer faithfulness, citation accuracy, access-policy compliance, and failure handling. Interaction logs can reveal unmet information needs, but they also require privacy protection and should not be treated as automatic ground truth for continual learning.

3.2 Multimodal foundation models for subsurface interpretation

Exploration integrates seismic volumes, well logs, core and thin-section images, maps, and text at different scales. Vision and multimodal foundation models can connect those sources

through promptable segmentation, cross-modal alignment, and retrieval. The SAG model, for example, fine-tuned a vision foundation model with LoRA for multiple geobody types. It accepts point, box, mask, and well-log prompts, propagates coherent interpretations from two-dimensional sections to three-dimensional (3D) volumes, and adapts across surveys and previously unseen geobodies (Gao et al., 2026). Core-image workflows similarly combine annotated thin-section, scanning-electron-microscopy, and computed-tomography images to identify minerals, pores, and fractures (Boruah and Ganapathi, 2015; Voorn et al., 2015; Luo et al., 2025; Liu et al., 2026). These capabilities turn interpretation into an iterative human-model dialogue in which an interpreter can add geological priors and refine a result instead of repeating complete manual delineation (Yu and Ma, 2021).

Petroleum foundation models are also being adapted to curve and field-specific tasks. Kunlun LLM has been applied to strike-slip-fault recognition and carbonate fracture-cavity prediction within three-dimensional seismic workflows (China National Petroleum Corporation, 2025). Gaia was evaluated on data from 250 wells. Its adapter tuning updated about 1/70 of the original parameters, reduced the mean absolute error by at least 20% against the reported deep-learning and large-model baselines, and improved blind-well generalization (Chen et al., 2025). At the organizational level, GeoFedNet evaluated structure classification, salt detection, and facies segmentation across four datasets; it remained within 1%-3% of centralized performance and outperformed isolated local models by as much as 15% (Ul Islam and Wali, 2026). Together, these cases demonstrate how domain transfer and privacy-preserving collaboration can support an exploration platform. Fig. 1 separates four functions within a multimodal petroleum system and illustrates how they form a supervised improvement loop. Domain adaptation updates the model vocabulary and task behavior with petroleum terms such as formation names, logging abbreviations, drilling events, and completion procedures. Cross-modal alignment links a seismic horizon, a well-log interval, a core image, and its geological description so that evidence referring to the same object can be compared (Wu et al., 2025). At the same time, structured retrieval joins document passages with well, depth, time, and model-version metadata, allowing an engineer to retrieve the correct analogue. Expert review checks fault continuity, facies plausibility, pressure or integrity constraints, and source provenance before approved corrections enter the training or retrieval corpus. These functions allow component-level testing and prevent the language model from being credited with the numerical performance of downstream interpretation models.

3.3 Agentic orchestration and hybrid physics-AI workflows

LLM-based agents translate an engineering objective into a sequence of data queries, model runs, checks, and reports. In a petroleum workflow, an agent may retrieve offset-well information, configure a trajectory or reservoir simulator, call an optimization routine, compare alternatives, and ask for

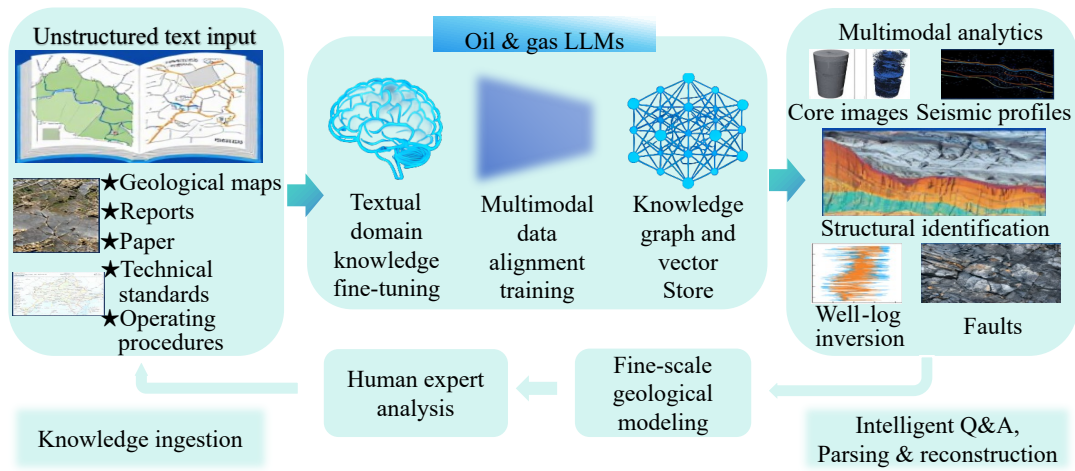


Fig. 1. Conceptual architecture for multimodal petroleum foundation-model systems.

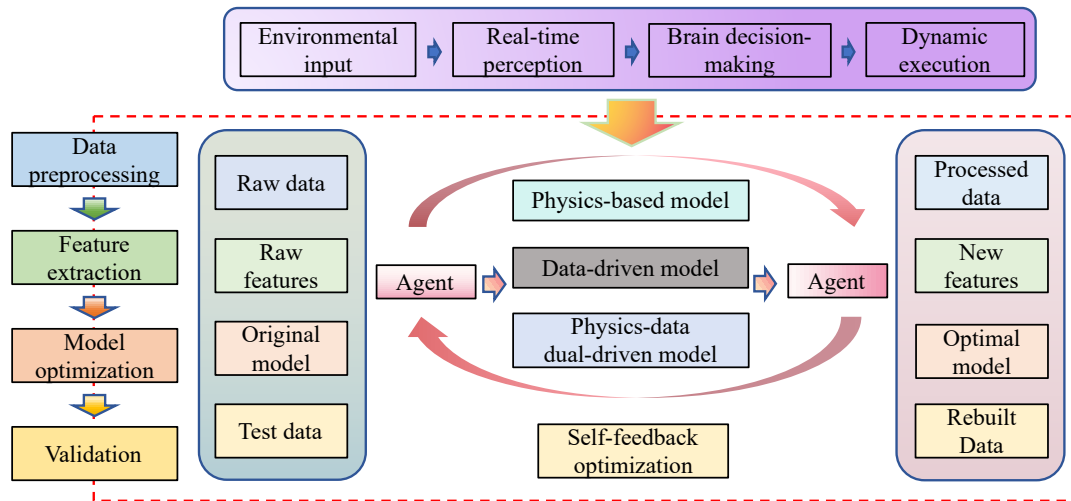


Fig. 2. Agentic optimization architecture.

approval at designated gates (Pan et al., 2024). The resulting system is beneficial because it links fragmented tools and data through a common semantic interface. Validated simulators, numerical optimizers, and task-specific machine-learning models remain the computational engines, while the LLM manages intent, context, and workflow state. As shown in Fig. 2, an LLM or planner coordinates governed data access and validated numerical tools, while physics-based and task-specific models act as execution units. As for the drilling design, the agent can retrieve offset-well constraints, call a trajectory or hydraulics model, compare feasible paths, and request approval before a plan is released. For reservoir management, it can configure history-matching or injection-allocation runs, pass the approved parameters to a simulator, and summarize uncertainty without generating the physical solution itself. A digital twin supplies a synchronized asset state; physics-based and data-driven models evaluate scenarios; an optimizer searches the feasible space; and the agent explains alternatives in engineering terms (Cui et al., 2025). When provenance, permissions, model versions, and operator decisions are recorded, the workflow becomes reusable orga-

nizational knowledge rather than an isolated analysis. This architecture provides the technical basis for the application scenarios reviewed below.

4. Application scenarios across the upstream life cycle

4.1 Oil and gas geology and exploration

Oil and gas exploration begins with an incomplete, multiscale description of the subsurface. The main application scenarios for foundation-model systems are seismic geobody interpretation, core and rock-image analysis, well-log reconstruction and petrophysical interpretation, geological knowledge retrieval, and model-assisted geosteering. A typical workflow aligns seismic attributes, well curves, images, and textual interpretations within a shared project context. Subsequently, modality-specific models generate structure, facies, or property candidates, while the language layer retrieves analogues and assembles an interpretation report (Syaifullah, 2019). This arrangement allows experts to move rapidly between regional geological concepts, local observations, and alternative inter-

pretations.

Recent examples of large language platforms add specific details to this scenario. For instance, SAG supports the interactive delineation of channels, salt bodies, paleokarst, and strata across surveys by combining a seismic image encoder with multimodal prompts and sequential prompt propagation for 3D volumes (Gao et al., 2026). Gaia represents well logs as sequences and transfers a large pre-trained model through adapter tuning. It reconstructs missing curves from spatial and sequential information learned from 250 wells, with at least a 20% reduction in MAE and better blind-well performance in the reported comparisons (Chen et al., 2025). At the field-platform scale, the Kunlun architecture has been applied to strike-slip-fault and fracture-cavity interpretation within enterprise data environments. The application focus is therefore not a single universal model but rather a coordinated interpretation environment spanning seismic, well, image, and document information.

In terms of the outputs of interpretation, they feed geological modeling and well placement. Fault surfaces, facies masks, reconstructed curves, and core-derived pore or fracture statistics can become soft or hard constraints for inversion, geostatistical modeling, and uncertainty ensembles. An agent can prepare inputs, launch alternative models, compare realizations against well control, and summarize how each interpretation changes volumetrics or trajectory risk. This scenario also creates a feedback loop: expert corrections become curated annotations, while blind-well and cross-survey tests reveal where a model needs local adaptation. The practical outcome is faster model iteration and a more consistent documentation of geological reasoning rather than replacing geological judgment.

4.2 Drilling and completion

Drilling and completion convert a geological target into a safe, productive well. The application scenarios of LLMs in this field integrate high-frequency surface and downhole data streams, daily reports, three-dimensional earth models, equipment constraints, and operating procedures. Specialist AI models predict or classify local conditions, while an LLM-enabled agent provides a conversational interface for planning, parameter recommendation, risk response, and post-job review (Li et al., 2022).

4.2.1 Intelligent drilling design, geosteering, and parameter optimization

In drilling operations, an LLM-enabled copilot can integrate offset-well experience, geological targets, engineering constraints, equipment limits, and real-time logging data to support planning and execution. This workflow is well demonstrated by Sinopec's drilling-rate optimization agent, integrated with the Great Wall model, by combining engineering calculations with task-specific AI models and considering 23 factors grouped into four categories: surface equipment capability, wellbore flow, drill-string motion, and rock-breaking mechanisms. With the maximum rate of penetration and minimum mechanical specific energy as objectives, the

system dynamically recommends drilling parameters, and thus reportedly achieved more than 80% accuracy in predicting rate of penetration 30 m ahead of the bit (Sinopec News, 2025a, 2025b). Similar platforms, such as Lumi and Delfi, extend this approach to seismic and well interpretation, directional drilling, and geosteering, enabling trajectory updates as the earth model evolves (China National Petroleum Corporation News, 2024). As illustrated in Fig. 3, the LLM mainly performs intent interpretation, information retrieval, workflow orchestration, and traceable explanation, while specialist modules remain responsible for rate-of-penetration prediction, hydraulics, torque-and-drag, and mechanical-specific-energy calculations. This division clarifies the respective roles of LLM-generated recommendations and engineering-model outputs within the drilling workflow (Hegde and Gray, 2017; Sinopec News, 2025a).

4.2.2 Early warning and risk control under complex drilling conditions

Lost circulation, kicks, wellbore collapse, and stuck pipes are preceded by combinations of pressure, flow, torque, drag, mud, cuttings, and formation signals that are difficult to interpret in isolation. An intelligent warning workflow continuously screens more than 20 parameter classes with time-series and anomaly models, retrieves similar events, and presents probable causes together with response procedures. Sinopec reported a geology-engineering coupled system with less than 10 s data latency and more than 85% warning accuracy for loss, kick, collapse, and sticking. In the HY18-2HF deep-shale-gas well in the Hongxing block, the system identified a loss-prone interval 210 m ahead; discussion and treatment were completed in approximately 2 h compared with about one day for the conventional process (Sinopec News, 2025a). This shows how early warning, procedural retrieval, and human coordination can shift risk management from event response to anticipatory control.

4.2.3 Intelligent fracturing optimization and decision support

Hydraulic-fracturing design requires the joint optimization of perforation and cluster layout, fluid and proppant systems, pump schedules, geomechanics, offset-well behavior, production targets, and economics. Conventional task-specific AI has already been applied to this process. Using data from about 6,000 wells across approximately 20 Western Siberian fields, a CatBoost model trained on 92 geological, well and treatment variables and 16 production descriptors achieved an R^2 of 0.815 and was subsequently used for design screening (Morozov et al., 2020). A successive study combined ridge regression, CatBoost, offset-well retrieval, and constrained gradient-free optimization using more than 5,000 wells and 6,687 fracturing operations, and field validation on 21 wells demonstrated its potential for treatment recommendation (Duplyakov et al., 2022). As illustrated in Fig. 4, this design workflow can be extended to real-time execution by integrating treatment design, pump-state monitoring, anomaly detection, equipment-risk prediction, and operator decisions. AI-R FRAC, for example, integrates decision support,

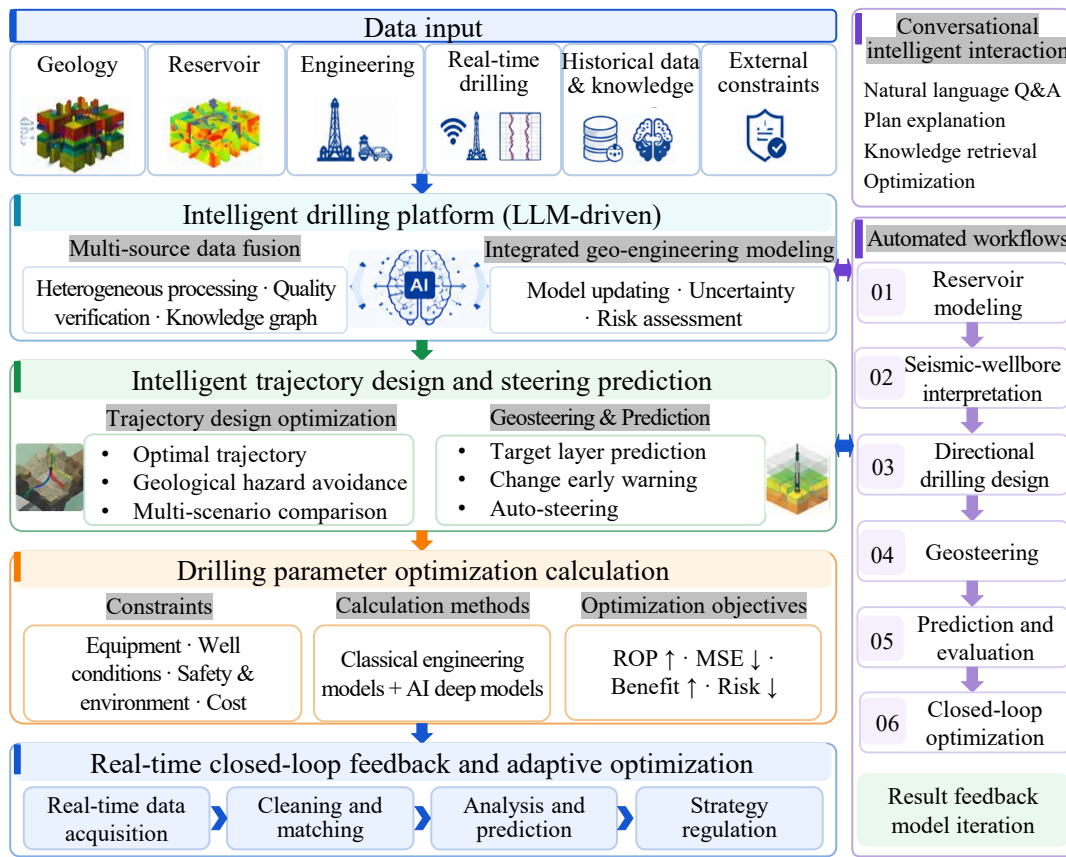


Fig. 3. Intelligent drilling design and optimization workflow.

equipment control, health monitoring, and job visualization through eight modules, with reported performance including 100% pressure-anomaly recognition, 97.8% plunger-pump risk-prediction accuracy, more than 36% improvement in site efficiency, and a 40% reduction in personnel (China Energy Net, 2025). Similarly, Sinopec's intelligent fracturing brain integrates geological and engineering knowledge with real-time treatment data (China Chemical Industry News, 2026). Together, these cases illustrate a common architecture in which LLM-enabled functions support information retrieval, context assembly, workflow coordination, and explanation, while numerical prediction, optimization, equipment control, and safety interlocks remain handled by domain-specific modules (Duplyakov et al., 2022; China Energy Net, 2025).

4.2.4 Intelligent completion design and economic evaluation

Completion design requires an integrated comparison of open-hole, cased-and-perforated, multistage-fractured, sand-control, inflow-control, and artificial-lift options. First, a scenario-oriented assistant extracts reservoir, wellbore, integrity, intervention, and cost variables from drilling and completion reports and retrieves analogue wells. Then, it configures inflow, wellbore, reliability, and economic calculations (Mitra, 2024; Li et al., 2025). LLM-assisted information extraction can convert legacy completion narratives into structured variables for well-integrity and failure-mode anal-

ysis, while engineering tools simulate inflow, pressure loss, sand-control performance, and intervention exposure. Finally, the alternatives are ranked by production profile, operability, failure risk, life-cycle cost, return on investment, and net present value. The deliverable is an auditable comparison matrix with sensitivities, assumptions, and a recommended design, allowing technical feasibility and economic value to be reviewed in the same workspace.

4.3 Oil and gas reservoir engineering

Reservoir engineering converts a geological description and production history into development decisions. The application scenarios include natural-language access to asset data, automated performance diagnostics, digital-twin surveillance, history-matching assistance, production forecasting, and the optimization of well controls or infill plans. The model stack combines databases and document stores with material-balance calculations, well-test and decline-curve analyses, simulators, surrogate models, and optimization tools (Zhang and Sun, 2021; Khalili and Ahmadi, 2023; Xue et al., 2023; He et al., 2024).

4.3.1 Reservoir performance analysis

A reservoir-performance assistant can translate a question such as “compare the production response of wells 1 and 2 over the past year and explain the difference” into governed database queries. It can then retrieve well events and tests,

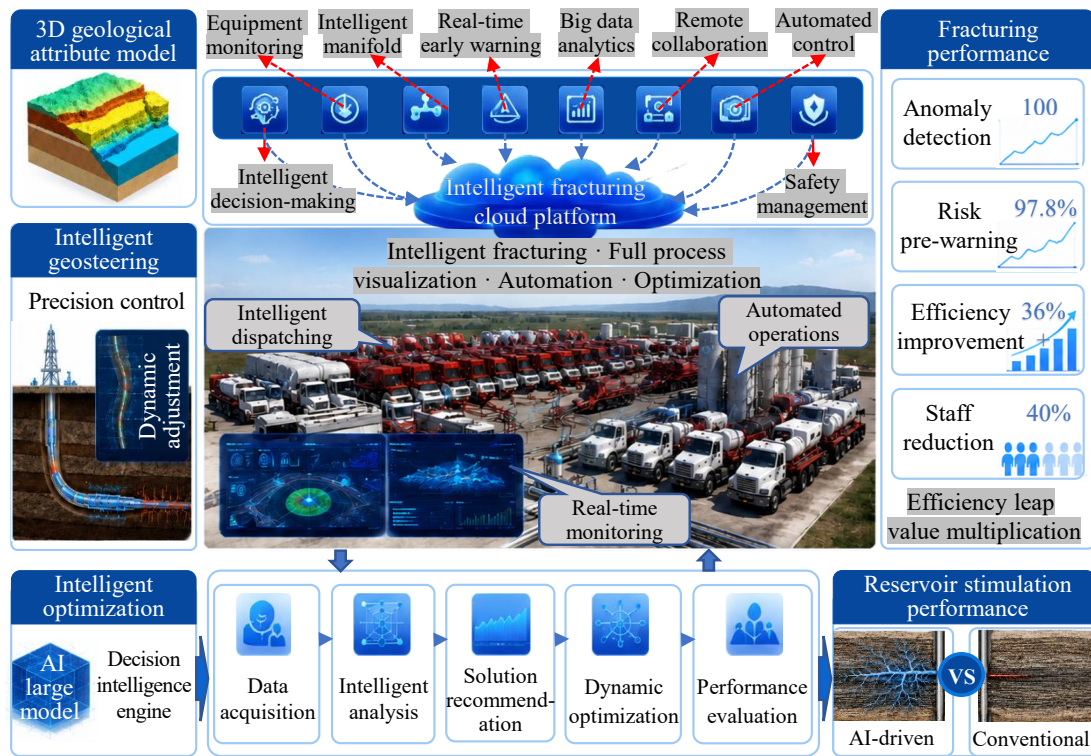


Fig. 4. Integrated geosteering and fracturing decision workflow.

plot dynamic curves, calculate water-cut and pressure indicators, and invoke connectivity analyses. The Daqing Oilfield and China University of Geosciences (Wuhan) system used Chinese-Alpaca-2.0 as its base, expanded the specialist vocabulary, and performed incremental pre-training on approximately 8.5 GB of reservoir-engineering text. Subsequently, instruction tuning and coupling with functional subsystems added named-entity recognition, text-to-SQL, tool invocation, curve plotting, indicator calculation, and preliminary injection-production connectivity analysis (Pan et al., 2024).

The above scenario changes reservoir analysis from navigation among databases and software menus to a dialogue centered on an engineering question. The LLM recognizes the asset, wells, interval, time window, and requested indicators, approved tools perform the calculations, and the answer combines plots, source records, and a technical narrative. The same interface can retain expert corrections and approved analytical templates, supporting consistent monthly surveillance and faster handover between asset teams (Singh et al., 2023; Zhao et al., 2025).

4.3.2 Reservoir dynamic analysis, digital twins, and decision support

A reservoir digital twin integrates a versioned geological model with production data, pressure and well-test observations, intervention history, state estimation, and forecasting models (Rasheed et al., 2020; Tosta et al., 2023). As new production responses become available, the twin updates the reservoir state, reconciles simulated and observed behavior, and evaluates alternatives such as water-injection reallocation,

shutoff, workover, and infill drilling. As illustrated in Fig. 5, this closed-loop workflow combines state synchronization, model updating, scenario evaluation, optimization, and engineering approval. Surrogate models can accelerate repeated scenario screening; for example, a multi-fidelity transfer-learning approach using a U-Net-type surrogate reduced computational cost by about 75% without sacrificing prediction accuracy and supported field-scale production optimization with time-varying well controls (Cui et al., 2025; Zhu et al., 2026). Within this architecture, the LLM layer supports workflow configuration, trade-off comparison, information retrieval, and decision-record generation, while validated simulators, surrogate models, and optimizers remain responsible for quantitative prediction and physical consistency (Zhang and Sun, 2021; Cui et al., 2025).

4.3.3 Production forecasting and development-strategy optimization

Production forecasting scenarios combine decline analysis, mechanistic flow models, machine learning, and operational-event context (Fan and Lai, 2025). For electric-submersible-pump wells, the reviewed hybrid framework first calculates theoretical multiphase production from lift and wellbore-flow mechanisms and then learns the residual error from historical production and equipment variables. The reported field example achieved more than 95% fitting accuracy for daily liquid production and a mean prediction error below 0.5 t (Du et al., 2025). Accordingly, forecasts can be evaluated under injection changes, choke or pump settings, shut-in timing, workovers, or infill development. The LLM contributes by ex-

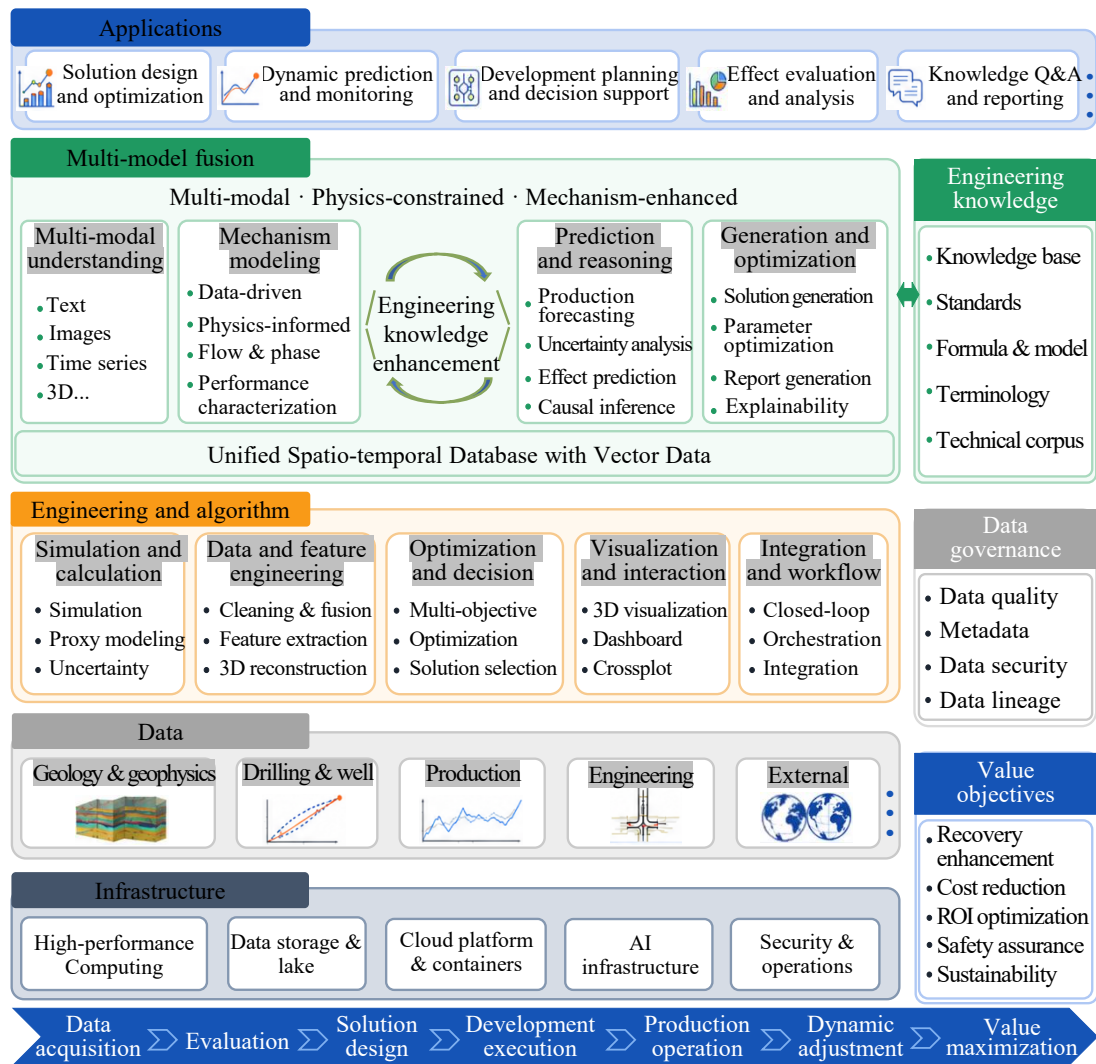


Fig. 5. Digital-twin framework for reservoir surveillance and development optimization.

tracting events from unstructured reports, selecting the relevant forecast configuration, and explaining why scenarios diverge, while the reservoir and optimization models produce the quantitative trajectories (Hu et al., 2020; Wang et al., 2024).

4.4 Production optimization

Production optimization operates on a continuously changing surface-subsurface system. The scenario therefore spans real-time sensing, production allocation, artificial-lift diagnosis, equipment-health prediction, work-order generation, process optimization, and economic scheduling. A layered platform connects supervisory control and data acquisition systems and data historians with mechanistic and data-driven models, while an LLM-enabled interface summarizes asset state and coordinates approved analyses (Oliveira et al., 2023; Zhang et al., 2025b).

4.4.1 Intelligent oil and gas field production management systems

A representative engineering example is provided by the intelligent oil-production management system reported by Sui et

al., 2025 for electric-submersible-pump wells in an East China Sea oilfield. The system integrates data processing, automated monitoring and proactive analysis to identify high-risk wells, diagnose gas interference, pump blockage, and abnormal temperature, pressure, current, or operating conditions, and recommends parameter adjustments or maintenance actions across different time horizons. More generally, as summarized in Fig. 6, intelligent production management can be organized into three layers: a data layer for cleaning and synchronizing real-time production, sensor, and maintenance data; a service layer that integrates warning, diagnosis, forecasting, and optimization models through APIs and microservices; and an application layer for visualization, decision support, and work-order management. By integrating multiple specialist models into a unified production-management framework, such systems can improve operational coordination and traceability, with recommendations linked to their source data, model version, threshold, and operator response (Sui et al., 2025; Zhang et al., 2025b).

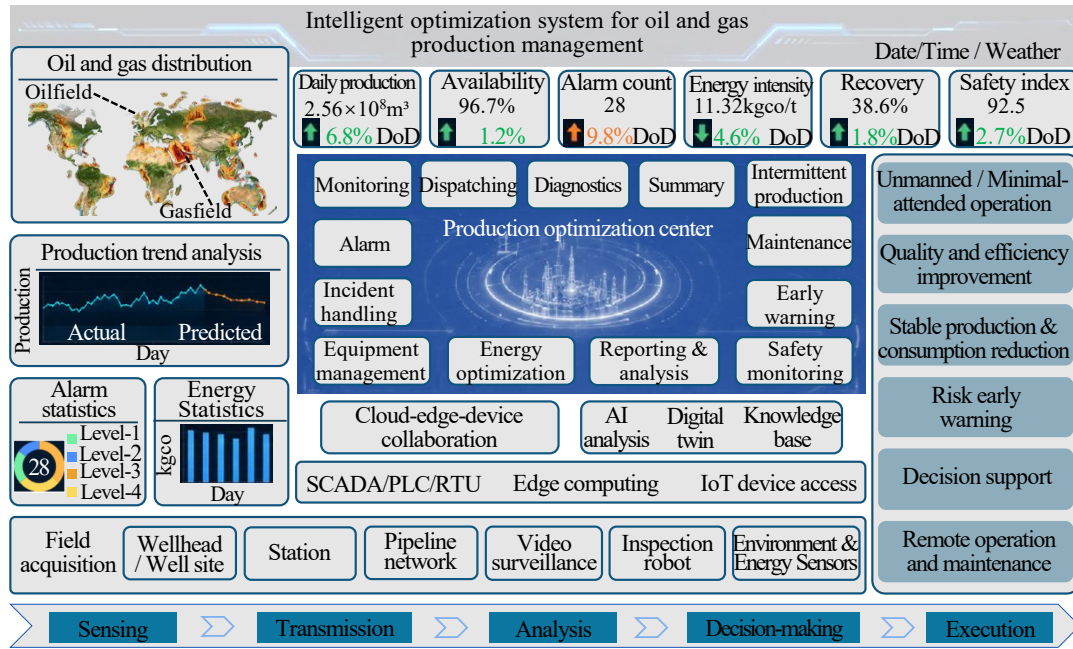


Fig. 6. Layered architecture for intelligent production management.

4.4.2 Production visualization, optimization and predictive maintenance

Digital twins extend production management by synchronizing models of physical assets with sensor, maintenance and operating data, allowing engineers to test a change before applying it to the field (Wanasinghe et al., 2020; Lin et al., 2024; Meza et al., 2024). The scale of this approach is well illustrated by the BP-Palantir collaboration. Palantir reported that its software integrated dynamic asset models with real-time data from more than two million sensors. A five-year agreement then added Artificial Intelligence Platform capabilities for analyzing data and recommending actions (Palantir Technologies, 2024). In practice, the operator can ask why a compressor, well, or processing train is deviating; the platform retrieves comparable events, runs diagnostic models, and displays likely causes and response options. The same environment supports predictive maintenance, injection reallocation, gas-lift optimization, and automated work-order drafting. As actions move closer to closed-loop execution, predefined operating envelopes, control-system interlocks, and operator authorization provide the transition from recommendation to bounded automation. Table 3 summarizes representative examples and separately records source type and deployment status.

5. Technical challenges, evaluation requirements and deployment controls

5.1 Data-related challenges

(1) Data ambiguity and uncertainty. Exploration data are inherently uncertain because the subsurface is complex, observations are sparse, and geophysical inverse problems are non-unique. Models trained directly on noisy data may learn

spurious correlations and produce biased predictions. Noise-aware training, physical regularization and probabilistic outputs can help represent uncertainty and allow experts to review or recalibrate predictions. For continuous streams such as logging-while-drilling and online production data, LLM-based agents could connect data acquisition, model updating and decision support in a monitored feedback loop (Abdar et al., 2021).

(2) Scarcity of high-quality labeled data. Geological interpretation depends heavily on expert annotation, which is costly and time-consuming, while core and other high-value samples are intrinsically limited. Meanwhile, few-shot learning can reduce, but not eliminate, the need for high-quality labeled examples. Useful approaches include self-supervised learning on unlabeled data, data augmentation, and semi-automated annotation. Cross-area transfer may generate candidate labels for seismic, well-log, and core data in a target area, after which experts can focus their effort on cases with high predictive uncertainty.

(3) Integration of heterogeneous multisource data. Oil and gas workflows combine seismic volumes, well logs, drilling data, production time series, and geological reports in different formats and at different scales. Because conventional LLMs are primarily text-based, non-text encoding and cross-modal alignment remain major challenges. One possible strategy is to encode waveforms, curves, images, and time series as tokens linked to spatial and temporal coordinates. Contrastive pre-training can then align seismic amplitudes, well-log responses, and descriptions of the same geological object in a shared representation space. The value of such representations depends on preserving physical meaning, scale, uncertainty, and provenance.

(4) Data silos and privacy. Industrial datasets are distributed across departments and companies and are often subject to

Table 3. Representative upstream applications, system boundaries and deployment statuses.

Scenario	Model or system	Boundary	Task	Deployment
Knowledge service	Sheng Xiaoli; MetaBrain	Demonstrated language-centered petroleum LLM	Search, drafting, cross-system access	Industrial use reported
Seismic interpretation	SAG vision foundation model	Demonstrated multimodal foundation model	Promptable 3D geobody delineation	Research validation
Well-log reconstruction	Gaia adapter-tuned model	Demonstrated large pre-trained model with adapter tuning	Missing-curve reconstruction	Research validation
Drilling design	Great Wall model+task-specific AI	Industrial agentic platform with mixed model attribution	ROP prediction and parameter recommendation	Field report
Drilling warning	Time-series/anomaly models+retrieval	Industrial AI warning platform; LLM role not isolated	Loss, kick, collapse, and sticking warning	Field application reported
Fracturing design	CatBoost+regression+optimizer	Conventional ML/optimization enabler; not an LLM	Forward prediction and inverse design	Field-tested workflow
Smart fracturing	AI-R FRAC platform	Industrial AI platform; language-model role not demonstrated	Anomaly recognition and equipment risk	Industrial platform
Reservoir analysis	Chinese-Alpaca-2.0+tools	Demonstrated LLM-enabled reservoir assistant	Text-to-SQL, plotting, connectivity analysis	Prototype/application
Reservoir optimization	U-Net surrogate+transfer learning	Industrial AI/digital-twin platform; LLM contribution not isolated	Digital-twin scenario evaluation	Research validation
Production management	Layered AI production platform	Conventional specialist AI platform; not a demonstrated LLM	ESP surveillance, diagnosis, and actions	Field application

confidentiality restrictions. However, federated learning allows organizations to train a shared model without exchanging raw data. In GeoFedNet, local seismic models were aggregated across heterogeneous clients for structure classification, salt detection, and facies segmentation. Performance remained within 1%-3% of centralized training and exceeded isolated local models by as much as 15% (Ul Islam and Wali, 2026). Secure aggregation, access control, and audit trails are still required when this pattern is transferred to proprietary field data.

5.2 Integration of domain knowledge and physical mechanisms

(1) Representation of domain knowledge. General-purpose LLMs have limited exposure to specialist terminology, field experience, and tacit operational knowledge. However, domain corpora, knowledge graphs, RAG, and domain-adaptive training can improve the relevance and traceability of model outputs. Automated collection and entity alignment may accelerate corpus construction, but source quality and licensing must be controlled. Importantly, retrieval systems should return evidence and provenance with each answer, while experts review uncertain or safety-critical content.

(2) Integration of physics and data-driven models. Drilling and reservoir behavior are constrained by conservation laws and constitutive relationships, whereas purely data-driven models can produce physically inconsistent predictions. Physics-informed neural networks incorporate governing-equation residuals or constraints, including conservation of mass, momentum and energy, into the training objective (Zhang et al., 2024; Kang et al., 2025). More broadly, LLM-enabled agents can invoke validated mechanistic simulators

and coordinate explicit physical-consistency checks. The language model orchestrates these operations, and it does not supply the governing physics or the numerical solution. These hybrid strategies may improve extrapolation and interpretability (Raissi et al., 2019; Karniadakis et al., 2021; Shan et al., 2023; Xie et al., 2023).

5.3 Model performance, evaluation and reliability

(1) Interpretability, traceability and evaluation. Petroleum LLMs should be evaluated at both the task and system levels. Knowledge services require retrieval recall, answer faithfulness, citation accuracy, and abstention tests. Agents require tool-selection accuracy, parameter validation, execution success, recovery from tool failure, and complete logs. Meanwhile, numerical predictions require calibration, uncertainty quantification, and out-of-distribution testing of the underlying models (Cao et al., 2026). However, human review remains necessary for safety-critical recommendations, but reviewer agreement and override behavior should also be recorded (Rudin, 2019).

(2) Generalization and few-shot adaptation. Models deployed in new blocks or unconventional reservoirs may fail when local geology differs from the training distribution (Saadat et al., 2024). At the same time, transfer learning, parameter-efficient fine-tuning, few-shot learning, and knowledge distillation can support adaptation with limited local data. Pre-training across geological, engineering, and textual sources may encode broadly applicable priors, while LoRA or AdaLoRA can incorporate a small target-area dataset. Distillation can reduce model size and inference latency for deployment.

(3) Real-time performance, safety and security. Latency requirements vary by workflow and must be derived from the control or decision process rather than asserted as a universal threshold. Safety-critical functions require deterministic fallbacks, bounded tool permissions, operator override, and separation between advisory output and control execution. Cybersecurity measures must protect model endpoints, retrieval stores, credentials, data interfaces, and control-system connections from prompt injection, data poisoning and unauthorized action.

5.4 Deployment and economic viability

(1) Computing and cost constraints. Training and serving LLMs require substantial computing resources, whereas edge hardware is limited. Pruning, quantization and distillation can reduce memory use and inference cost. Cloud-edge architectures may allocate computationally demanding tasks to centralized resources while retaining latency-sensitive processing on site.

(2) Dynamic systems and continuous learning. Reservoir and production conditions evolve over time, causing data distributions and model performance to drift. Online or periodic updating can help models track these changes, while updates must be versioned, validated and monitored to prevent performance degradation or the introduction of unsafe behavior (De Lange et al., 2022).

(3) Evaluation of economic value. Intelligent systems may require substantial investment, while their benefits may be distributed across productivity, downtime, safety, and knowledge retention. Evaluation frameworks should therefore relate total life-cycle cost to measurable operational outcomes while quantifying uncertainty. Therefore, demonstrated value under representative field conditions is essential before large-scale deployment.

6. Conclusions and outlook

LLMs and foundation models are reshaping upstream oil and gas by connecting domain knowledge, operational data, engineering models, and decisions. Their role is not to replace petroleum science but to coordinate heterogeneous information and specialized tools. LLMs should support semantic interaction and workflow orchestration, while validated simulators remain responsible for prediction, optimization and control, with this layered architecture offering a more credible path than deploying isolated general-purpose models.

The decisive challenge is to provide dependable value under field conditions rather than model scale or benchmark performance. Systems must preserve geological meaning, physical consistency, provenance, cybersecurity, and human accountability while transferring across assets. Progress requires governed asset memory, hybrid model-simulator architectures with testable interfaces, and life-cycle governance from version control to field-based economic assessment. Evaluation should jointly consider technical accuracy, workflow reliability, operational risk, transferability, and economic value.

Deployment should proceed from scenario-specific copi-

lots, through cross-disciplinary agents, to bounded autonomy for well-instrumented tasks. Each stage should be evidence-based and retain deterministic safeguards and immediate human override. The strategic goal is, instead of an autonomous “AI petroleum engineer,” providing accountable, physics-grounded engineering intelligence that connects models, data, simulators, digital twins, and experts. Such infrastructure can accelerate decisions and reuse expertise, provided that adoption is justified by demonstrated safety, reliability and net value across the asset life cycle.

Acknowledgements

This work was supported by the Ningbo Top Talent Science and Technology Project (No. 2024DJRC010001), the National Natural Science Foundation of China (No. 52604053), the Zhejiang Provincial Natural Science Foundation of China (Nos. LQN26E040006 and LMS26E040005) and the Ningbo Yongjiang Talent Programme (No. 2025A-127-G).

Conflicts of interest

The authors declare no competing interest.

Open Access This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Abdar, M., Pourpanah, F., Hussain, S., et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 2021, 76: 243-297.
- [Anthropic. Introducing Claude Opus 4.6. 2026.](#)
- Boruah, A., Ganapathi, S. Microstructure and pore system analysis of Barren Measures shale of Raniganj field, India. *Journal of Natural Gas Science and Engineering*, 2015, 26: 427-437.
- Brown, T. B., Mann, B., Ryder, N., et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 2020, 33: 1877-1901.
- Cao, L., Jiang, F. J., Chen, Z. X., et al. A novel interpretable machine learning framework for predicting gas-bearing properties of tight sandstone reservoirs. *Petroleum Science*, 2026, 23(7): 3805-3833.
- Chen, Z., Zhang, Y., Li, J., et al. Artificial intelligence large model for logging curve reconstruction. *Petroleum Exploration and Development*, 2025, 52(3): 842-854.
- Cheng, D., Huang, S., Wei, F. Adapting large language models via reading comprehension. Paper Presented at the Twelfth International Conference on Learning Representations (ICLR 2024), Vienna, Austria, 7-11 May, 2024.
- [China Chemical Industry News. Sinopec's intelligent fracturing brain goes online. 2026. \(in Chinese\)](#)
- [China Energy Net. The world's first systematic smart fracturing solution AI-R FRAC is unveiled. 2025. \(in Chinese\)](#)
- China National Petroleum Corporation. 2024 Annual Report. Beijing, China, 2025.
- [China National Petroleum Corporation News. Where is oil and](#)

gas AI heading? 2024. (in Chinese)

- Chowdhery, A., Narang, S., Devlin, J., et al. PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 2023, 24(240): 1-113.
- Cui, J. W., Sun, W. Y., Jeong, H., et al. Efficient deep-learning-based surrogate model for reservoir production optimization using transfer learning and multi-fidelity data. *Petroleum Science*, 2025, 22(4): 1736-1756.
- De Lange, M., Aljundi, R., Masana, M., et al. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7): 3366-3385.
- Devlin, J., Chang, M. W., Lee, K., et al. BERT: Pre-training of deep bidirectional transformers for language understanding. Paper Presented at Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- Di, S., Cheng, S., Cao, N., et al. AI-based geo-engineering integration in unconventional oil and gas. *Journal of King Saud University-Science*, 2021, 33(6): 101542.
- Dongo, D. F., Relvas, S. Evaluating the role of the oil and gas industry in energy transition in oil-producing countries: A systematic literature review. *Energy Research & Social Science*, 2025, 120: 103905.
- Duan, H., Ma, C., Wang, Z., et al. Large language model of oil and gas cognition constructed and applied in Shengli Oil-field. *Petroleum Science and Technology Forum*, 2024, 43(6): 46-55. (in Chinese)
- Du, C., Gong, F., Zhou, Y., et al. A mechanistic-based data-driven modeling framework for predicting production of electric submersible pump wells in offshore oilfield. *Geoenergy Science and Engineering*, 2025, 246: 213603.
- Duplyakov, V. M., Morozov, A. D., Popkov, D. O., et al. Data-driven model for hydraulic fracturing design optimization. Part II: Inverse problem. *Journal of Petroleum Science and Engineering*, 2022, 208: 109303.
- Fan, N., Lai, F. Patch-Transformer modeling for production forecasting in tight gas reservoirs: A case study of the Yan'an field. *Energy & Fuels*, 2025, 39(1): 432-443.
- Gao, H., Wu, X., Liang, L., et al. A foundation model empowered by a multi-modal prompt engine for universal seismic geobody interpretation across surveys. *Information Fusion*, 2026, 125: 103437.
- Goodfellow, I., Bengio, Y., Courville, A. *Deep Learning*. Cambridge, Cambridge, USA, MIT Press, 2016.
- Guo, D., Yang, D., Zhang, H., et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 2025, 645(8081): 633-638.
- Hammad, K. J., Al-Turki, A. A., Sudirman, S. B., et al. Enhancing reservoir model history matching with AI surrogate and ensemble iterative algorithms. Paper SPE 221028 Presented at SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 23-25 September, 2024.
- Han, T., Adams, L. C., Bressemer, K. K., et al. Comparative analysis of multimodal large language model performance on clinical vignette questions. *JAMA*, 2024, 331(15): 1320-1321.
- Haouel, C., Nemeslaki, A. Digital transformation in oil and gas industry: Opportunities and challenges. *Periodica Polytechnica Social and Management Sciences*, 2024, 32(1): 1-16.
- Hassani, H., Silva, E. S. Big data: A big opportunity for the petroleum and petrochemical industry. *OPEC Energy Review*, 2018, 42(1): 74-89.
- He, Y., Wei, B., Zhao, J., et al. Accurate determination of nano-confined minimum miscible pressure to aid CO₂ enhanced oil recovery and storage in unconventional reservoirs. *Advances in Geo-Energy Research*, 2024, 12(2): 141-155.
- Hegde, C., Gray, K. E. Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *Journal of Natural Gas Science and Engineering*, 2017, 40: 327-335.
- Hu, Z., Li, Y., Chang, J., et al. New model for production prediction of shale gas wells. *Energy & Fuels*, 2020, 34(12): 16486-16492.
- Kang, M. L., Zhou, J., Zhang, J., et al. An integrated method of data-driven and mechanism models for formation evaluation with logs. *Petroleum Science*, 2025, 22(3): 1110-1124.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., et al. Physics-informed machine learning. *Nature Reviews Physics*, 2021, 3(6): 422-440.
- Khalili, Y., Ahmadi, M. Reservoir modeling and simulation: Advancements, challenges, and future perspectives. *Journal of Chemical and Petroleum Engineering*, 2023, 57(2): 343-364.
- Kirillov, A., Mintun, E., Ravi, N., et al. Segment anything. Paper Presented at Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2-6 October, 2023.
- Koroteev, D., Tekic, Z. Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. *Energy and AI*, 2021, 3: 100041.
- Kuang, L., Liu, H., Ren, Y., et al. Application and development trend of artificial intelligence in petroleum exploration and development. *Petroleum Exploration and Development*, 2021, 48(1): 1-14.
- LeCun, Y., Bengio, Y., Hinton, G. Deep learning. *Nature*, 2015, 521(7553): 436-444.
- Lewis, P., Perez, E., Piktus, A., et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 2020, 33: 9459-9474.
- Li, X., Bai, L., Zhao, Y., et al. Review of the application of artificial intelligence and big data technology in oilfield completion design and optimization: A systematic analysis from parameter optimization to safety management. *Advances in Resources Research*, 2025, 5(1): 31-45.
- Li, G., Song, X., Tian, S., et al. Intelligent drilling and completion: A review. *Engineering*, 2022, 18: 33-48.
- Li, J. X., Zhang, T. C., Zhu, Y. R., et al. Artificial general intelligence for the upstream geoenergy industry: A review. *Gas Science and Engineering*, 2024, 131: 205469.

- Lin, B., Jin, Y., Cao, Q., et al. Developing a large language model for oil- and gas-related rock mechanics: Progress and challenges. *Natural Gas Industry B*, 2025, 12(2): 110-122.
- Lin, B., Zhu, H., Jin, Y., et al. Modeling approach and case studies of digital twin in drilling and production of oil and gas fields. *Petroleum Science Bulletin*, 2024, 9(2): 282-296. (in Chinese)
- Liu, H., Ren, Y. L., Li, X., et al. Research status and application of artificial intelligence large models in the oil and gas industry. *Petroleum Exploration and Development*, 2024, 51(4): 1049-1065.
- Liu, C., Shan, L., Zhao, J., et al. Deep learning in digital rock technology: A comprehensive review. *Advanced IntelliEngineering*, 2026, 1(1): 4-20.
- Luo, X., Sun, J. M., Chi, P., et al. ThinGPT: Describing sedimentary rock thin section images with a multimodal large language model. *Petroleum Science*, 2025, 22(12): 5020-5033.
- Manning, C. D., Schütze, H. *Foundations of Statistical Natural Language Processing*. Cambridge, Cambridge, USA, MIT Press, 1999.
- Meza, E. B. M., de Souza, D. G. B., Copetti, A., et al. Tools, technologies and frameworks for digital twins in the oil and gas industry: An in-depth analysis. *Sensors*, 2024, 24(19): 6457.
- Mirza, M. A., Ghoroori, M., Chen, Z. Intelligent petroleum engineering. *Engineering*, 2022, 18: 27-32.
- Mitra, N. K. *AI and Digital Technology for Oil and Gas Fields*. Boca Raton, USA, CRC Press, 2024.
- Mohammadpoor, M., Torabi, F. Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*, 2020, 6(4): 321-328.
- Morozov, A. D., Popkov, D. O., Duplyakov, V. M., et al. Data-driven model for hydraulic fracturing design optimization: Focus on building digital database and production forecast. *Journal of Petroleum Science and Engineering*, 2020, 194: 107504.
- Motaei, E., Ganat, T. Smart proxy models art and future directions in the oil and gas industry: A review. *Geoenery Science and Engineering*, 2023, 227: 211918.
- Nguyen, T., Gosine, R. G., Warrian, P. A systematic review of big data analytics for oil and gas industry 4.0. *IEEE Access*, 2020, 8: 61183-61201.
- Oliveira, R. G., Almeida, L. F., Lazo Lazo, J. G., et al. Smart oil field management system using evolutionary intelligence. *IEEE Access*, 2023, 11: 45798-45814.
- OpenAI. *ChatGPT: Optimizing language models for dialogue*. OpenAI, 2022.
- OpenAI. *Learning to reason with LLMs*. OpenAI, 2024.
- OpenAI. *GPT-5 system card*. OpenAI, 2025.
- OpenAI. *Introducing GPT-5.4*. OpenAI, 2026a.
- OpenAI. *GPT-5.6: Frontier intelligence that scales with your ambition*. OpenAI, 2026b.
- OpenAI. *GPT-6 Astra: A new generation of intelligence*. OpenAI, 2026c.
- Ouyang, L., Wu, J., Jiang, X., et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 2022, 35: 27730-27744.
- Palantir Technologies. *Palantir and bp agree to five-year strategic relationship with new AI capabilities*. 2024.
- Pan, H., Liu, J., Gong, B., et al. Construction and preliminary application of large language model for reservoir performance analysis. *Petroleum Exploration and Development*, 2024, 51(5): 1357-1366.
- Radford, A., Kim, J. W., Hallacy, C., et al. Learning transferable visual models from natural language supervision. Paper Presented at Proceedings of the 38th International Conference on Machine Learning, Vienna, Austria, 18-24 July, 2021.
- Radford, A., Wu, J., Child, R., et al. Language models are unsupervised multitask learners. *OpenAI Blog*, 2019, 1(8): 9.
- Raffel, C., Shazeer, N., Roberts, A., et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020, 21(140): 1-67.
- 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, 378: 686-707.
- Rasheed, A., San, O., Kvamsdal, T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 2020, 8: 21980-22012.
- Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 2019, 1(5): 206-215.
- Saadat, M., Hashemi, H., Nabi-Bidhendi, M. Generalizable data driven full waveform inversion for complex structures and severe topographies. *Petroleum Science*, 2024, 21(6): 4025-4033.
- Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks*, 2015, 61: 85-117.
- Singh, A., Jia, T., Nalagatla, V. Generative AI enabled conversational chatbot for drilling and production analytics. Paper SPE 216267-MS Presented at Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 2-5 October, 2023.
- Shan, L., Liu, C., Liu, Y., et al. Physics-informed machine learning for solving partial differential equations in porous media. *Advances in Geo-Energy Research*, 2023, 8(1): 37-44.
- Sircar, A., Yadav, K., Rayavarapu, K., et al. Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 2021, 6(4): 379-391.
- Sinopec News. *Intelligent drilling: Artificial intelligence leads the transformation of oil and gas engineering technology*. 2025a. (in Chinese)
- Sinopec News. *Large models and big data create dual intelligent engines for the oil and gas industry*. 2025b. (in Chinese)
- Sinopec News. *Oil and gas giants are using AI to reduce costs and boost efficiency*. 2025c. (in Chinese)

- Sui, X., Lu, X., Ji, Y., et al. Intelligent oil production management system based on artificial intelligence technology. *Processes*, 2025, 13(1): 133.
- Syaifulallah, K.I. The use of big data in the oil and gas upstream industry: A comparison between Norway and Indonesia. *Indonesian Journal of Energy*, 2019, 2(1): 14-28.
- Tariq, Z., Aljawad, M. S., Hasan, A., et al. A systematic review of data science and machine learning applications to the oil and gas industry. *Journal of Petroleum Exploration and Production Technology*, 2021, 11(12): 4339-4374.
- Tosta, M., Oliveira, G. P., Wang, B., et al. APyCE: A Python module for parsing and visualizing 3D reservoir digital twin models. *Advances in Geo-Energy Research*, 2023, 8(3): 206-210.
- Ul Islam, M. S., Wali, A. GeoFedNet: Federated learning for privacy-aware, robust, and generalizable seismic interpretation. *Computers & Geosciences*, 2026, 207: 106060.
- Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- Voorn, M., Exner, U., Barnhoorn, A., et al. Porosity, permeability and 3D fracture network characterisation of dolomite reservoir rock samples. *Journal of Petroleum Science and Engineering*, 2015, 127: 270-285.
- Wanasinghe, T. R., Wroblewski, L., Petersen, B. K., et al. Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. *IEEE Access*, 2020, 8: 104175-104197.
- Wang, J., Luo, X., Zhang, X., et al. Artificial general intelligence (AGI) applications and prospect in oil and gas reservoir development. *Processes*, 2025, 13(5): 1413.
- Wang, T., Wei, Q., Xiong, W., et al. Current status and prospects of artificial intelligence technology application in oil and gas field development. *ACS Omega*, 2024, 9(3): 3173-3183.
- Weizenbaum, J. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 1966, 9(1): 36-45.
- Wu, J., Luo, R. Z., Luo, L., et al. LLMs-guided parameters prediction of tight sandstone reservoirs. *Petroleum Science*, 2025, 22(12): 5005-5019.
- Xie, C., Du, S., Wang, J., et al. Intelligent modeling with physics-informed machine learning for petroleum engineering problems. *Advances in Geo-Energy Research*, 2023, 8(2): 71-75.
- Xue, L., Li, D., Dou, H. Artificial intelligence methods for oil and gas reservoir development: Current progresses and perspectives. *Advances in Geo-Energy Research*, 2023, 10(1): 65-70.
- Yin, S., Fu, C., Zhao, S., et al. A survey on multimodal large language models. *National Science Review*, 2024, 11(12): 403.
- Yu, S., Ma, J. Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 2021, 59(3): e2021RG000742.
- Zhang, J., Braga-Neto, U., Gildin, E. Physics-informed neural networks for multiphase flow in porous media considering dual shocks and interphase solubility. *Energy & Fuels*, 2024, 38(18): 17781-17795.
- Zhang, F., Feng, Y., Li, X. Application of big data and artificial intelligence in oilfield production optimization and intelligent management: A systematic analysis of real-time monitoring and parameter optimization. *Advances in Resources Research*, 2025a, 5(1): 16-30.
- Zhang, H., Xu, J., Cui, H., et al. When geoscience meets foundation models: Toward a general geoscience artificial intelligence system. *IEEE Geoscience and Remote Sensing Magazine*, 2025b, 13(1): 2-41.
- Zhang, S., Chen, W. China's energy transition pathway in a carbon neutral vision. *Engineering*, 2022, 14: 64-76.
- Zhang, T., Sun, S. An exploratory multi-scale framework to reservoir digital twin. *Advances in Geo-Energy Research*, 2021, 5(3): 239-251.
- Zhang, Y. Analysis of China's energy efficiency and influencing factors under carbon peaking and carbon neutrality goals. *Journal of Cleaner Production*, 2022, 370: 133604.
- Zhang, Y., Pan, Y., Zhong, T., et al. Potential of multimodal large language models for data mining of medical images and free-text reports. *Meta-Radiology*, 2024, 2(4): 100103.
- Zhao, X., Wei, S., Wang, L., et al. Large language model empowered automated reservoir agent: A win-win strategy for reservoir intelligent management and transformation. Paper SPE 229646-MS Presented at Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 3-6 November, 2025.
- Zhu, J., Dai, J., Xie, M., et al. Machine learning-assisted injection-production optimization for offshore low-permeability reservoirs. *Energy & Fuels*, 2026, 40(7): 3525-3537.
- Zhu, J., Du, Y., Zhang, W., et al. AI-driven knowledge management in oil and gas: A large language model approach to operational excellence. Paper SPE 230403 Presented at SPE Annual Caspian Technical Conference and Exhibition, Baku, Azerbaijan, 25-27 November, 2025.