energies

MDPI

Review

From Data to Decision-Making: A Review of Large Language Model Applications Across the Full Lifecycle of Geothermal Resource Development

Xiaoguang Li [1,2], Zewei Shang [1], Tianrun Zhao [1], Shiji Wang [1] and Ziwang Yu [3,*]

[1] College of Civil Engineering, Changchun Institute of Technology, Changchun 130012, China
[2] Jilin Province Key Laboratory for Civil Engineering Safety and Resilience, Changchun 130012, China
[3] College of Construction Engineering, Jilin University, Changchun 130012, China
* Correspondence: yuziw@jlu.edu.cn

Abstract

Geothermal resource development involves heterogeneous geological and engineering records, uncertain reservoir parameters, and decisions constrained by physical processes and safety requirements. This narrative and critical review examines how large language models (LLMs), including retrieval-augmented and tool-assisted systems, could support knowledge-intensive tasks throughout geothermal exploration, development, and operation. It distinguishes direct geothermal evidence from empirical evaluations in adjacent fields and from conceptual proposals; results from petroleum engineering, mineral exploration, and broader energy systems are not treated as proof of geothermal performance. Peer-reviewed studies demonstrate useful capabilities in technical document extraction, domain question answering, knowledge-graph reasoning, and interactions with scientific software, but direct geothermal engineering validation remains limited. For reservoir inversion, drilling-risk warning, and design optimization, quantitative assessment must rely on measured data and validated physical or numerical models, with LLMs acting as auditable interfaces and information-integration tools. The review proposes a task-specific benchmark framework and examines cross-domain transfer, multimodal alignment, knowledge-graph and retrieval collaboration, uncertainty quantification, and LLM–physical-model coupling. These frameworks are literature-informed proposals rather than implemented geothermal systems. Key priorities include independent geothermal test sets, expert-annotated reference answers, calibrated uncertainty, physics-based verification, and transparent reporting of model failure. The resulting synthesis identifies credible current applications and the evidence requirements for reliable future deployment.

Keywords: geothermal resources; Large Language Models (LLMs); Retrieval-Augmented Generation (RAG); intelligent decision-making

Check for updates

Academic Editor: Boris Igor Palella

Received: 24 September 2026
Revised: 18 October 2026
Accepted: 21 October 2026
Published: 23 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Geothermal energy, as a low-carbon and dispatchable renewable energy source, has become an increasingly important component of the global energy transition and sustainable development owing to its capability for continuous energy supply with minimal dependence on climatic conditions [1]. According to resource depth and temperature, geothermal resources can be broadly classified into shallow, medium-deep, and deep high-temperature systems, which are primarily applied to building heating and cooling, district energy supply, and large-scale power generation, respectively [2]. With the advancement

of global carbon neutrality initiatives and the expansion of geothermal development, increasing demands have been placed on resource assessment, reservoir characterization, and engineering decision-making.

Despite these advances, significant challenges remain throughout the geothermal development lifecycle. From a data perspective, large volumes of unstructured information—including geological reports, drilling records, reservoir evaluation documents, and operational reports—are continuously generated during exploration, drilling, reservoir characterization, and system operation, yet their effective utilization remains limited. From the perspectives of reservoir characterization and numerical modeling, geothermal systems involve complex Thermal–Hydraulic–Mechanical–Chemical (THMC) coupled processes, where substantial parameter uncertainty and the high computational cost of conventional numerical simulations hinder efficient analysis [3]. From a decision-making perspective, critical tasks such as resource assessment, drilling design, and operational optimization continue to rely heavily on expert knowledge, while systematic mechanisms for knowledge reuse and quantitative decision support remain insufficient. Consequently, the integration of multi-source information, intelligent knowledge extraction, and data-driven decision support has become a pressing requirement for the sustainable development of the geothermal industry.

Recent advances in Large Language Models (LLMs) have demonstrated remarkable capabilities in natural language understanding, knowledge representation, and complex reasoning, providing a promising avenue for addressing these challenges. Through techniques such as domain-specific fine-tuning, Retrieval-Augmented Generation (RAG), and multimodal learning, LLMs can effectively integrate domain knowledge with diverse unstructured data sources, exhibiting considerable potential in knowledge extraction, information integration, intelligent question answering, and engineering decision support [4]. For the geothermal sector, which has traditionally relied on geological interpretation and expert judgment, LLMs have the potential to serve as an intelligent bridge connecting data, domain knowledge, and engineering decision-making.

Recent subsurface modeling studies reinforce the importance of geological heterogeneity and uncertainty in AI-assisted decision-making. Teng et al. (2025) developed a zonal-permeability, coupled CO_2-enhanced coalbed–methane model for abandoned mine goafs, illustrating spatially variable permeability and coupled transport [5]. More directly in the geothermal field, Li et al. (2026) reviewed machine-learning-based multiscale reservoir modeling, surrogate models, and digital twins, noting opportunities for uncertainty-aware forecasting alongside limitations under geological and operational shifts [6]. These studies highlight the distinction between uncertainty in retrieved language-model outputs and uncertainty quantified by physical or predictive models.

Against this background, this review examines the research progress and potential roles of LLMs throughout geothermal resource exploration, development, and utilization. The major contributions are summarized as follows:

(1) Clarifying the role of LLMs in geothermal engineering by distinguishing their applicability and complementary relationships with conventional machine learning, natural language processing, knowledge graphs, and physics-based models.
(2) Systematically reviewing key enabling technologies, including RAG, domain-specific fine-tuning, and multimodal learning, and summarizing the principal approaches for integrating LLMs with geothermal domain knowledge.
(3) Summarizing representative application scenarios of LLMs across geothermal exploration, reservoir characterization, drilling and field development, Enhanced Geothermal Systems (EGS) operation and management, and knowledge services, while an-

alyzing the correspondence between different model architectures and engineering tasks.

(4) Identifying current challenges and future research directions, including data quality, model reliability, multimodal integration, and engineering deployment, and proposing a roadmap that integrates domain-specific foundation models, physics-informed constraints, multi-agent collaboration, and digital twin technologies.

Literature Selection and Evidence Assessment

This study adopts a narrative and critical review approach. The initial literature search identified relevant publications through Web of Science Core Collection, Scopus, ScienceDirect, and Google Scholar from 2018 to 2025, using combinations of "Large Language Models," "Generative Artificial Intelligence," "Foundation Models," "Geothermal Energy," "Enhanced Geothermal Systems," "Geothermal Reservoir," "Geothermal Exploration," and "Geothermal Engineering." Selected peer-reviewed studies published in 2026 were subsequently incorporated through targeted updates to cover relevant methodological developments. Because direct studies of geothermal LLM applications remain limited, work from petroleum engineering, geoscience, and broader energy systems was considered where it offered relevant methods or comparative evidence; those source-domain results were not treated as independent proof of applicability to geothermal engineering.

The literature was synthesized across technical foundations, representative application tasks, and engineering limitations. For the study-level comparison presented in Section 4.1, emphasis was placed on peer-reviewed journal studies with identifiable LLM functions, application domains, evaluated tasks, data sources, and reported validation procedures or results. Research on conventional machine learning, geological data infrastructure, and numerical simulation was considered supporting technical context, not direct empirical evidence of LLM performance. The evidence was distinguished descriptively as (A) direct empirical validation in geothermal tasks, (B) empirical evaluation in an adjacent field without independent geothermal validation, or (C) conceptual or proposed applications lacking target-task validation. The six journal studies included in this comparison fall into category B; the three categories describe the location and type of evidence, not its quality or engineering maturity.

For each representative study, the reported evidence was examined in terms of its tested task, system configuration, dataset characteristics, evaluation measures, comparisons, and validation setting. Methodological considerations included data representativeness, potential training–test leakage, evaluation design, independence of human or model-based judging, and completeness of reporting. Because the studies use different datasets, tasks, and performance measures, numerical results are presented in their original contexts. The resulting comparison summarizes evaluation evidence, application boundaries, and geothermal-specific validation requirements. Cross-domain transferability was examined in relation to data distribution, technical terminology, governing physical processes, and site-specific engineering constraints; Section 4.2 develops these dimensions.

The remainder of this paper is organized as follows. Section 2 introduces the core technologies of LLMs and their adaptation to geothermal applications. Section 3 reviews representative applications of LLMs throughout the geothermal development lifecycle. Section 4 discusses the major challenges and future research directions. Finally, Section 5 presents the conclusions and perspectives for future research.

2. The Geothermal Development Lifecycle and the Core Technologies and Integration Foundations of Large Language Models

2.1. Current Status of the Geothermal Industry and Its Technology Framework

2.1.1. Current Status of the Geothermal Industry

Geothermal energy is a clean, low-carbon, and dispatchable renewable energy source capable of providing a continuous supply of heat and electricity regardless of climatic conditions or diurnal cycles. Owing to these characteristics, it has become an important component of baseload energy in the global energy transition [7]. By 2024, the global installed geothermal power generation capacity had reached approximately 15.1 GW, with the United States (approximately 3.97 GW), Indonesia (approximately 2.1 GW), and the Philippines (approximately 1.9 GW) ranking as the leading producers. In recent years, the global installed capacity has maintained an annual growth rate of approximately 2–3%. In comparison, the direct utilization of geothermal energy has achieved a substantially larger scale, with an annual energy use of approximately 245 TWh, supporting a wide range of applications, including district heating, greenhouse agriculture, and industrial heat supply [8].

For the purpose of discussing development pathways, geothermal resources can be broadly grouped into shallow, medium-deep, and deep systems, although the boundaries vary with geological and national classification conventions [2]. An alternative classification based on the thermodynamic exergy of geothermal fluids has also been proposed [9]. The corresponding depth ranges, temperature intervals, key utilization technologies, and representative application scenarios are illustrated in Figure 1 Among these, shallow geothermal systems represent the most mature utilization pathway, with Ground Source Heat Pumps (GSHPs) serving as the dominant technology. Increasingly, GSHPs are being integrated with underground thermal energy storage technologies, making them the mainstream solution for building heating and cooling applications. Medium-deep geothermal resources constitute the primary heat source for direct heating and district energy supply and have already been deployed on a commercial scale in countries such as Germany and France. Deep high-temperature geothermal resources are primarily utilized for electricity generation, with flash steam power plants and binary cycle power plants representing the most mature technologies, typically achieving power generation efficiencies of 10–17% [10]. Among these technologies, Enhanced Geothermal Systems (EGS) overcome the limitations imposed by naturally permeable reservoirs through artificial reservoir stimulation, thereby substantially expanding the exploitable geothermal resource base. Demonstration projects have been implemented in North America and Europe, and EGS is widely regarded as one of the most promising technological pathways for the future expansion of the geothermal industry [11].

Despite its considerable development potential, geothermal exploitation continues to face inherent challenges, particularly the high costs associated with exploration and drilling and the substantial uncertainty in reservoir characterization [11]. Supported by global low-carbon policies and international technological collaboration, emerging geothermal markets in Southeast Asia and Africa have experienced rapid growth, with newly installed capacities increasing by approximately 200–300 MW per year, thereby becoming new drivers of industrial expansion. In the long term, the coordinated development of shallow, medium-deep, and deep geothermal resources, together with the integration of geothermal energy with photovoltaic, wind power, and energy storage systems, is expected to further improve overall system efficiency and economic value [12]. Such integrated energy systems will strengthen the role of geothermal energy as a reliable baseload resource within the future global energy mix.

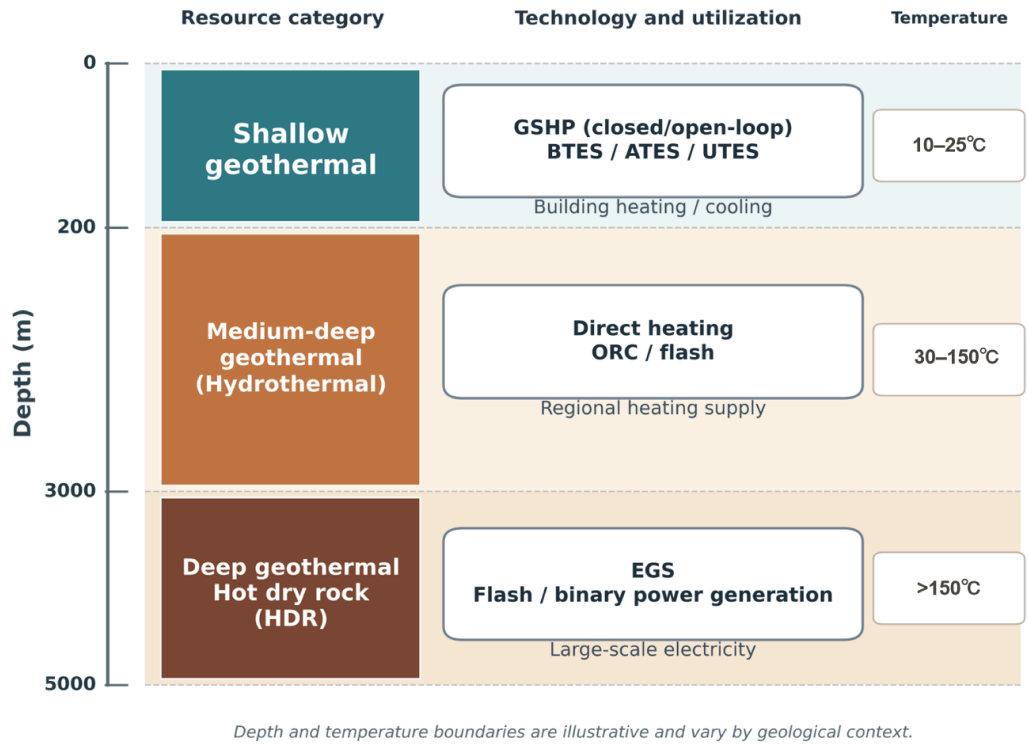


Figure 1. Overview of shallow, medium-deep, and deep geothermal resource categories, typical depth and temperature ranges, and representative utilization technologies.

2.1.2. Core Technology Framework for Geothermal Development

The technological framework for geothermal development is closely associated with resource depth. Geothermal systems at different depths exhibit distinct technological pathways, engineering challenges, and technical bottlenecks.

Shallow geothermal systems are generally classified into closed-loop and open-loop configurations, which differ substantially in heat transfer mechanisms, wellfield layouts, and operational strategies (Figure 2) [13,14]. The principal technical challenge lies in accurately characterizing the heterogeneity of subsurface thermophysical properties. Moreover, long-term operation may lead to underground thermal imbalance, highlighting the need for refined system design and operational control through multi-source data integration.

Related studies of buried subsurface thermal infrastructure further emphasize the influence of soil thermal properties and heating regimes on subsurface temperature evolution. Ahmad et al. (2025) experimentally examined soil thermal responses around underground power cables under steady and cyclic heating conditions [15]. Ahmad et al. (2026) reviewed deterministic, data-driven, and hybrid approaches to heat-transfer modeling for buried subsea cables, highlighting sediment heterogeneity and transient thermal boundary conditions [16]. Although underground cables and geothermal systems differ in operating conditions and heat-exchange configurations, these studies provide complementary insights into subsurface heat-transfer processes.

Medium-deep geothermal systems typically employ paired production–reinjection wells for heat extraction. Low- and medium-temperature resources are primarily utilized for direct heating through surface heat exchangers, whereas medium- to high-temperature resources can be coupled with flash steam technology or an Organic Rankine Cycle (ORC) for electricity generation [17]. Their major technical bottlenecks arise from uncertainties in reservoir parameter characterization caused by geological heterogeneity and the high computational cost of multi-physics numerical simulations. Consequently, current tech-

nological advances mainly focus on extracting knowledge from heterogeneous geological data and improving the accuracy of reservoir parameter prediction.

Deep geothermal development is centered on Enhanced Geothermal Systems (EGSs), which rely on ultra-deep drilling and hydraulic stimulation to create artificial fracture networks. The implementation of EGSs requires overcoming a series of technical challenges, including drilling and completion under ultra-high temperature and pressure conditions, fracture network optimization, coupled Thermal–Hydraulic–Mechanical–Chemical (THMC) simulations, and microseismic monitoring [18]. The primary challenges stem from the strong uncertainties associated with both reservoir properties and engineering operations, together with the urgent demand for real-time decision-making during field deployment. Addressing these issues requires dynamic risk identification and operational control based on the integration of multi-source real-time data.

Overall, the three categories of geothermal development share three common technological bottlenecks. First, although large volumes of multi-source heterogeneous data are generated throughout the geothermal lifecycle, their utilization remains limited, and valuable information contained in geological reports and engineering documents has yet to be fully exploited. Second, current understanding of subsurface reservoirs lags behind engineering requirements, making geological uncertainty difficult to quantify accurately. Third, critical engineering decisions continue to rely heavily on expert knowledge, while systematic mechanisms for knowledge reuse and intelligent decision support remain underdeveloped. To address these challenges, a hierarchical intelligent framework is required. For shallow geothermal systems, the emphasis should be placed on multi-source data integration and refined system design; for medium-deep geothermal systems, priority should be given to domain knowledge extraction and reservoir parameter prediction; and for deep geothermal systems, intelligent processing of real-time data and risk-aware decision-making are of primary importance. Owing to their powerful capabilities in understanding unstructured information and integrating knowledge across heterogeneous data sources, LLMs provide a promising technological foundation for advancing intelligent geothermal development across all resource categories.

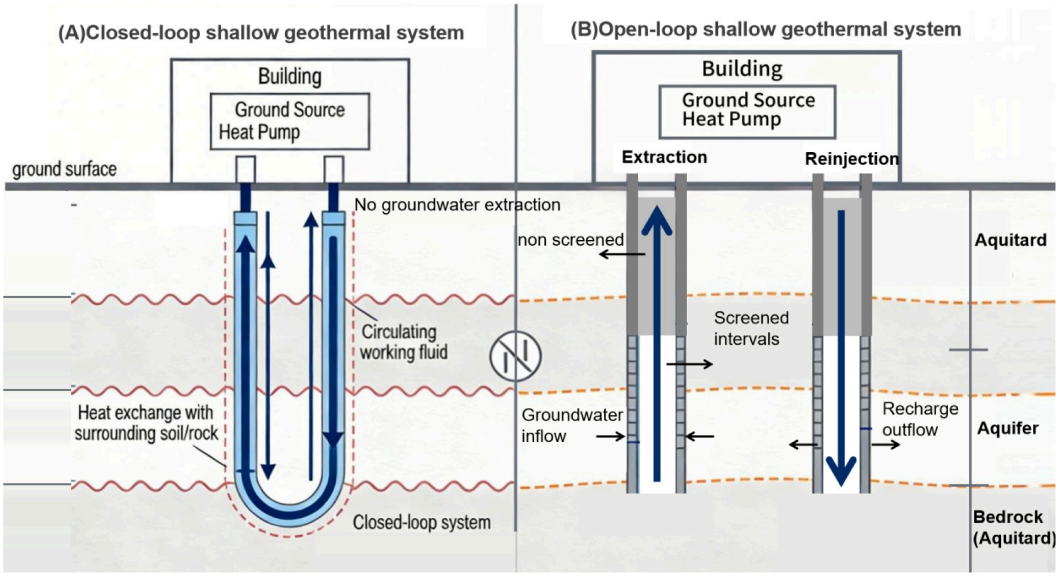


Figure 2. Schematic comparison of closed-loop and open-loop shallow geothermal systems (not to scale). Note: Blue arrows indicate fluid circulation, whereas thin arrows show local flow directions or identify labeled components.

2.2. Core Technologies of Large Language Models and Their Adaptation to Geothermal Applications

2.2.1. Technical Positioning and Architectural Framework of LLMs

The underlying architecture of Large Language Models (LLMs) is predominantly built upon the Transformer framework, which has evolved into three mainstream architectural paradigms: encoder-only, encoder–decoder, and decoder-only models. Among these, the decoder-only autoregressive architecture, represented by the GPT, LLaMA, and Qwen model families, has emerged as the dominant technical paradigm owing to its superior text generation capability, scalability, and generalization performance. Among the architectural approaches to efficient scaling, sparse Mixture-of-Experts (MoE) models activate only selected expert components for each input, as illustrated by Switch Transformers [19]. The overall architecture and key technological components of modern LLMs are illustrated in Figure 3.

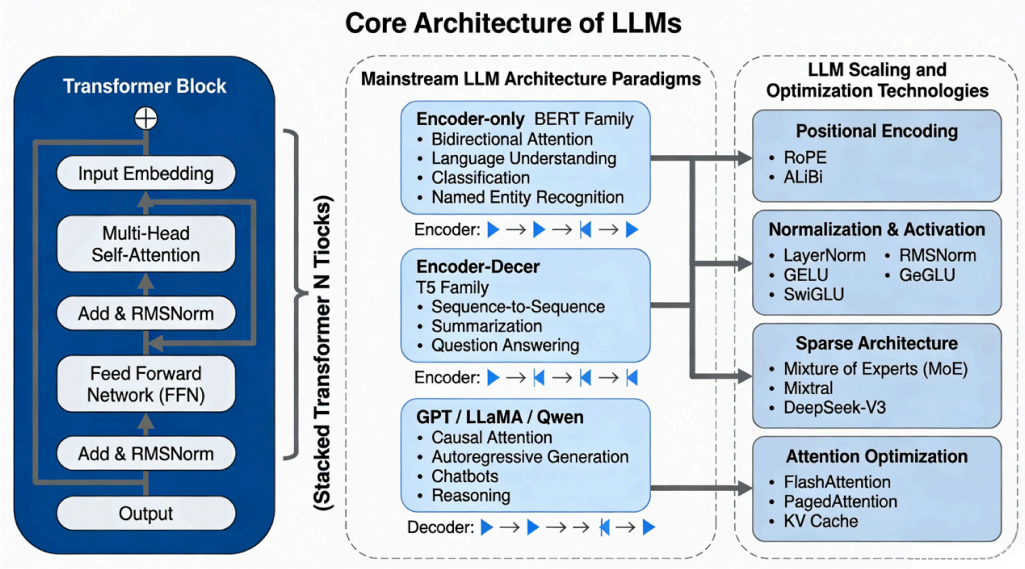


Figure 3. Core Transformer architecture, mainstream model paradigms, and key optimization technologies of modern large language models. Note: Arrows indicate information flow or connections between architectural components. Blue triangles schematically represent token-flow patterns.

It is important to distinguish Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), the Transformer architecture, foundation models, and LLMs, as they represent different levels within the AI technology hierarchy and should not be used interchangeably. AI is the broadest discipline concerned with the development of intelligent systems, whereas ML and DL constitute its principal methodological branches. The Transformer is a deep neural network architecture based on the self-attention mechanism and serves as the dominant backbone of contemporary foundation models. Foundation models are general-purpose models pretrained on massive datasets and subsequently adapted to a wide range of downstream tasks. LLMs represent a typical category of foundation models designed specifically for natural language processing, with most current implementations adopting the Transformer architecture. In essence, the Transformer defines the model architecture, foundation models represent the model paradigm, and LLMs constitute a specific model type, each occupying a distinct position within the AI technology stack [20].

From a technical perspective, LLMs exhibit both significant differences from and complementarity with conventional ML models and knowledge graphs (KGs). Compared with traditional ML and DL approaches, the primary innovation of LLMs lies not merely in improving predictive accuracy for individual tasks, but in establishing a unified paradigm of pretraining–transfer learning–instruction tuning, enabling a single model

to perform diverse tasks, including information extraction, knowledge-based question answering, text generation, and logical reasoning, while effectively integrating knowledge across multiple domains. Compared with KGs, LLMs demonstrate superior capabilities in natural language understanding and human–computer interaction, whereas KGs provide distinct advantages in structured knowledge representation, logical reasoning, and interpretability. Consequently, the two approaches are highly complementary in domain-specific applications.

LLMs generate text from learned statistical regularities rather than by directly solving physical governing equations, so their outputs alone cannot establish physical validity [20]. In contrast, published geothermal AI work has generally evaluated specialized prediction models rather than LLMs as numerical reservoir solvers [21]. In geothermal engineering, reservoir evolution is governed by complex coupled Thermal–Fluid–Mechanical (TFM) and Thermal–Fluid–Mechanical–Chemical (TFMC) (also referred to as Thermal–Hydraulic–Mechanical–Chemical (THMC)) processes. These processes are described by systems of coupled partial differential equations that must be solved numerically using specialized simulators, such as TOUGH2 and COMSOL Multiphysics, to satisfy fundamental physical conservation laws. General-purpose LLMs do not incorporate physics-based solution engines. Without deep integration with specialized techniques such as Physics-Informed Neural Networks (PINNs), they cannot independently solve complex partial differential equations or guarantee compliance with fundamental physical constraints, including mass and energy conservation. Consequently, LLMs cannot replace specialized numerical simulation tools in quantitatively intensive tasks such as reservoir parameter inversion, temperature field prediction, and coupled reservoir process simulation.

To improve their applicability in specialized domains such as geothermal engineering, LLM-based systems may be adapted through instruction tuning, Parameter-Efficient Fine-Tuning (PEFT), and retrieval-augmented generation (RAG). RAG connects generated text to retrieved documents, but retrieved passages may be incomplete, outdated, or inaccurate; the overall reliability still depends on retrieval quality and source verification [22]. In high-risk engineering scenarios such as drilling optimization, stimulation design, and operational risk assessment, retrieved information cannot substitute for site-specific safety checks. Human-in-the-loop (HITL) validation is therefore required for proposed engineering decisions, with domain specialists reviewing evidence, units, regulations, and physical constraints before implementation.

In summary, LLMs should be regarded as knowledge integration and decision-support tools rather than replacements for physics-based models, numerical simulators, or expert judgment in geothermal engineering. Their primary value lies in integrating scientific literature, engineering data, domain expertise, and numerical simulation results to provide comprehensive information support and analytical insights for complex engineering problems, while leaving quantitative physical analysis and final engineering decisions to specialized computational tools and domain experts.

2.2.2. Core Capabilities of LLMs and Their Applicability to the Geothermal Industry

The geothermal development lifecycle encompasses resource exploration, reservoir characterization, drilling and completion design, and operational management. Owing to its complex knowledge system, heterogeneous data sources, and strong demand for interdisciplinary collaboration, the geothermal industry is highly compatible with the core capabilities of LLMs. Their application value can be summarized in three major aspects.

First, LLMs possess strong capabilities for processing unstructured information. Over decades of geothermal development, vast amounts of unstructured data—including exploration reports, well-log interpretation reports, technical standards, and scientific

publications—have been accumulated. Conventional manual processing of these resources is labor-intensive and inefficient. LLMs can assist in literature screening, information extraction, synthesis, and document summarization, subject to verification against the original sources. As a complementary approach, knowledge-graph-extended RAG has been evaluated for question answering over structured relationships [23]. These capabilities provide an important technological foundation for the digital management of geothermal knowledge assets.

Second, LLMs demonstrate strong capabilities in cross-domain knowledge integration and decision support. Unlike conventional natural language processing models that are typically designed for specific tasks, LLMs can simultaneously integrate geological conditions, engineering parameters, economic indicators, and safety constraints to support site selection, engineering design, and operational optimization. Their principal strength lies in synthesizing heterogeneous knowledge, establishing logical relationships among diverse information sources, and providing interpretable decision support, rather than performing high-precision numerical calculations. Consequently, LLMs can effectively complement traditional engineering tools by enhancing the transparency and interpretability of engineering decision-making.

Third, multimodal foundation-model approaches offer a route to connect geological text with seismic images, well-log sequences, core photographs, and numerical simulation fields. However, jointly interpreting these modalities requires site-aware and depth/time-aware alignment, appropriate modality-specific encoders, and verification of physical meaning; concatenating heterogeneous files into a text prompt is not equivalent to multimodal reasoning. Results reported for geoscience foundation models illustrate the potential of seismic interpretation guided by well-log prompts [24], but geothermal-specific integration and operational reliability remain insufficiently validated. Therefore, multimodal LLMs should support—not replace—specialized geophysical processing, reservoir simulation, and expert interpretation.

Despite these advantages, the application boundaries of LLMs in geothermal engineering should be clearly defined. Quantitative tasks, including reservoir parameter inversion, resource assessment, and coupled thermo-hydraulic reservoir simulation, should continue to rely primarily on conventional ML models and specialized numerical simulators. For example, previous studies have demonstrated that Bayesian optimization-based ML regression models can accurately predict bottom-hole temperature (BHT) and static formation temperature (SFT), confirming the effectiveness and cost-efficiency of data-driven approaches for wellbore temperature estimation. The added value of LLMs lies in interpreting quantitative modeling results, integrating heterogeneous information, and facilitating human-centered decision-making, rather than directly performing rigorous physical calculations. Accordingly, LLMs are more appropriately regarded as an intelligent interaction layer that connects engineering data, domain knowledge, and specialized computational models. Moreover, any outputs related to engineering safety or critical design parameters should always undergo expert verification before practical implementation to ensure the reliability and safety of engineering decisions.

3. Applications of Large Language Models in the Geothermal Sector

3.1. Intelligent Question Answering and Knowledge Management for Geothermal Applications

3.1.1. Intelligent Question Answering Systems

Domain-specific question-answering (QA) offers an important knowledge-access application for geothermal engineering. Energy-domain model adaptation [25] and geoscientific knowledge-graph retrieval [26] illustrate complementary directions for specialist question answering. Multi-source energy-system RAG [27] and geological-record standard-

ization [28] address related information-processing tasks. Their evaluation datasets, tasks, and performance measures differ across the original application domains.

(1) Evidence-grounded QA with traceable domain sources

Geoscience knowledge services can combine authoritative source passages, specialist terminology, and source-traceable answer generation. For example, Zhou and Li (2025) examined a geoscience LLM coupled with knowledge-graph construction and lightweight RAG for geological question answering [26]. Its evaluation in mineral exploration illustrates an architecture potentially transferable to geothermal knowledge services.

In an evidence-grounded workflow, document identifiers, spatial or stratigraphic context, observation dates, and parameter units should accompany retrieved passages. Where relevant sources disagree or the available context is insufficient, the response should expose the conflict or explicitly defer the decision to expert review. Source-linked retrieval and expert escalation are important safeguards for geothermal knowledge services [23,26].

A geothermal QA benchmark would need independently annotated questions spanning exploration reports, borehole descriptions, reservoir properties, and engineering constraints. Its evaluation should include answer correctness, citation support, refusal behavior, and robustness to changed or missing source documents. Peer-reviewed experience from geological document standardization offers complementary evidence for testing controlled vocabulary and source provenance [28].

(2) Domain-adapted geological retrieval and information organization

Geological records contain specialized terms and inconsistent naming conventions, making retrieval quality and terminology normalization central design concerns. Zhou and Li (2025) explored geoscience knowledge-graph extraction with a lightweight RAG approach, whereas Yan et al. (2026) investigated RAG-supported standardization of geological borehole descriptions [26,28]. These studies evaluate complementary mechanisms for organizing and retrieving geological information in their respective source domains.

Yan et al. (2026) evaluated a geological database standardization workflow using 100 boreholes from 21 engineering projects in Macao, comprising 908 stratigraphic samples [28]. Their original setting was urban engineering geology. The transfer of the method to geothermal logs would require a separate expert-labelled test set covering geothermal lithologies, depth intervals, measurement units and inconsistent source terminology.

Evaluation of domain-oriented systems spans operational measures (e.g., query volume, response latency, and cost), retrieval or information-processing measures (e.g., ranking quality, entity-extraction F1, and formation standardization), and generated-answer measures (e.g., reference-answer accuracy and expert assessment). Because tasks, datasets, and protocols differ, cross-study performance values are descriptive rather than directly comparable [26,28].

(3) Configurable geoscience knowledge and reasoning pipelines

Domain-specific geological reasoning can combine adapted LLMs with retrieved documentation and structured knowledge relations. Ma et al. (2026) studied such a pipeline in mineral exploration, combining domain adaptation, RAG and knowledge-graph reasoning [29]. It illustrates a potential design pathway for other geological tasks, while its published results remain specific to mineral-exploration questions.

The integration of multiple information sources requires entity alignment, documented retrieval provenance, and temporal consistency. For geothermal workflows, useful test cases could involve borehole aliases, lithological classifications, field-measured temperatures, and site-specific operational constraints.

Accordingly, a configurable geothermal knowledge service should be validated on held-out wells and projects and should be assessed by independent domain experts, rather than relying on out-of-domain benchmark scores to establish engineering readiness.

(4) Foundation-model adaptation for related energy tasks

EnergyGPT, reported by Chebbi and Kolade (2026), adapts LLaMA 3.1-8B to the energy sector through full-parameter supervised fine-tuning and parameter-efficient LoRA training [25]. The journal study evaluates 476 energy-domain questions (233 multiple-choice, 100 true/false, and 143 open-ended). On multiple-choice questions, the base model achieved 73% accuracy versus 86% for EnergyGPT-SFT and 89% for EnergyGPT-LoRA. However, on true/false questions the base model achieved 95% and both adapted variants 94%, with no statistically significant improvement reported. These task-dependent results support domain adaptation for some forms of energy QA, not a universal performance gain or geothermal-specific accuracy.

Peer-reviewed energy-system RAG studies also support cross-source synthesis and traceable retrieval [27]. Applications in geothermal engineering additionally require consistent geological terminology, reservoir context, and appropriate engineering safety constraints.

Table 1 summarizes five representative energy- and geoscience-oriented LLM systems by architecture, source data, evaluation focus, and reported results. Because these studies address different tasks and use different datasets and assessment protocols, the comparison is descriptive rather than a numerical ranking.

Table 1. Representative LLM-based systems in energy and geoscience domains: evaluation scope and evidence.

System/ Framework	Technical Approach	Base Model	Knowledge Source Scale	Evaluation Focus	Reported Results and Scope	Typical Applications
Zhou and Li (2025) [26]	Geoscience LLM + KG construction + LightRAG	Geoscience language model	Mineral-exploration texts and entities	Entity extraction/geological QA	Mean entity-extraction F1 = 0.8353 in mineral geoscience; no geothermal benchmark	Mineral geological retrieval and QA
Yan et al. (2026) [28]	RAG-supported geological data standardization	LLM with geological retrieval	100 boreholes; 21 projects; 908 samples	Formation standardization; interpretability	Engineering-geology evaluation; no geothermal benchmark	Structured geological record processing
Ma et al. (2026) [29]	Domain adaptation + RAG + KG reasoning	Domain-adapted geological LLM	1201 expert-designed mineral questions	Domain QA accuracy; terminology coverage	Reported 89.99% mineral-domain QA accuracy	Specialist mineral exploration QA
EnergyGPT [25]	Energy-domain SFT and LoRA	LLaMA 3.1-8B	476 energy-domain questions	MCQ/T-F accuracy; human ratings	MCQ: 86% SFT, 89% LoRA vs. 73% base; T/F 94% vs. 95% base	Broader energy-domain QA

Table 1. *Cont.*

System/ Framework	Technical Approach	Base Model	Knowledge Source Scale	Evaluation Focus	Reported Results and Scope	Typical Applications
Arslan et al. (2024) [27]	Multi-source RAG–LLM for energy transitions	Domain-supported LLM + RAG	Multi-source energy system information	Application-based RAG evaluation	Published energy-system study; no geothermal benchmark	Multi-source energy knowledge support

Note: Operational efficiency, information-processing quality, and generated-answer accuracy assess different tasks and system properties. Reported values use study-specific datasets and evaluation protocols and are not directly comparable.

3.1.2. Intelligent Processing of Geological Literature and Engineering Documents

The geothermal sector has accumulated a vast amount of unstructured textual data, including scientific publications, exploration reports, drilling logs, technical specifications, and engineering documentation. LLMs provide powerful capabilities for the automated processing of these documents, enabling efficient information extraction, knowledge organization, and knowledge accumulation [30]. Building upon these capabilities, and in combination with knowledge graph (KG) construction and enhanced data infrastructure, LLMs can facilitate the intelligent management of geothermal geological data. The overall technical framework is illustrated in Figure 4.

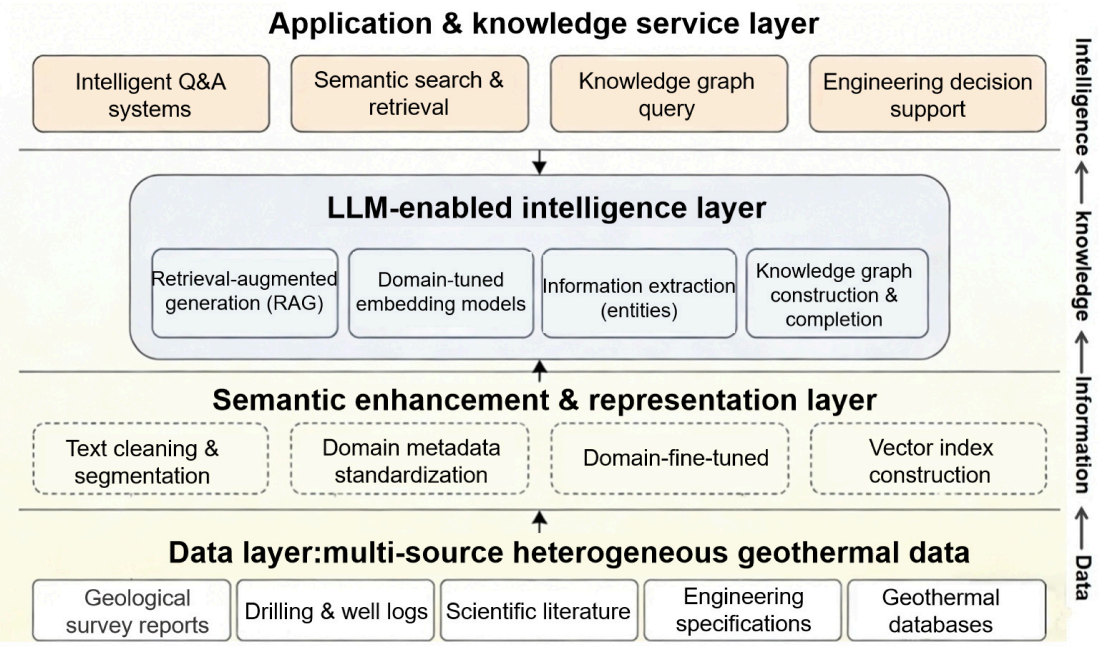


Figure 4. Overall framework of LLM-enabled intelligent management for geothermal geological data.

LLMs support the automated processing of geothermal documents through three progressively advanced capabilities: (i) automated literature screening for corpus construction, (ii) information extraction from engineering documents for knowledge acquisition, and (iii) semantic interpretation of technical standards and engineering cases to support regulatory compliance and engineering practice.

(1) Automated literature screening and corpus construction

Automated corpus construction requires identifying relevant technical literature, extracting structured information, and maintaining traceable links to the source documents.

Peer-reviewed studies have demonstrated that LLMs can extract structured information from scientific articles [30] and support the review of technical specification documents [31]. For geothermal literature organization, task-specific classification accuracy, corpus coverage, and data quality should be evaluated using labeled geothermal records.

(2) Information extraction from drilling logs and engineering documents

Information extraction from historical well records has been demonstrated in the oil and gas sector, where the distribution of geological settings, document formats, and reporting conventions may differ from geothermal drilling. Geothermal records can additionally emphasize high-temperature intervals, fracture-related circulation losses, fluid chemistry, and production–reinjection conditions. Therefore, extraction performance on petroleum documents cannot be assumed to transfer unchanged: geothermal-specific records and specialist terminology require independent validation.

Ma et al. (2024) used Llama 2 to extract location and depth information from 160 historical orphaned oil and gas well documents [32]. Their complete workflow achieved 100% accuracy on clean PDF records, whereas accuracy on unstructured image-based well records decreased to 70%. The contrast highlights the influence of document format and preprocessing on extraction performance. Evaluation on geothermal drilling reports should include scanned records, lithological descriptions, drilling-event terminology, and local reporting conventions.

(3) Semantic interpretation of technical standards and engineering case documents

Engineering standards and safety regulations are important sources of domain-specific constraints in geothermal project implementation. General-purpose studies of LLM-assisted technical-document review, such as Lee et al. (2024) [31], and well-record extraction by Ma et al. (2024) [32] demonstrate relevant text-processing capabilities, but neither validates automated compliance judgments for geothermal regulations. A standards-assistance workflow would need to retrieve the applicable jurisdiction-specific document and version, verify citations against the governing clauses, distinguish advisory text from mandatory requirements, and require qualified engineering review before any safety-related decision.

3.1.3. Automated Construction and Applications of Geothermal Knowledge Graphs

Knowledge graphs (KGs) represent domain entities, attributes, and relations in an explicit structure, while large language models (LLMs) support document interpretation and natural-language interaction. For geothermal applications, the two components are complementary but not interchangeable: a KG can encode well identifiers, stratigraphic relations, measurements, and provenance, whereas retrieval-augmented generation (RAG) can retrieve documentary evidence and synthesize source-linked answers. The following discussion distinguishes published adjacent-domain demonstrations from a proposed geothermal KG–RAG collaboration workflow [23,26].

(1) Knowledge extraction

LLMs can extract candidate entities and relations from geological reports, drilling logs, and technical specifications, but proposed triples should be checked against a defined ontology and the source passages from which they were derived. Zhou and Li (2025) demonstrated a domain-specific LLM coupled with LightRAG for ontology construction, entity and relation extraction, and geoscience question answering in mineral exploration [26]. This methodology provides a starting point for developing and evaluating geothermal knowledge graphs.

(2) Knowledge alignment and integration

Geothermal descriptions may use inconsistent names for identical wells, stratigraphic units, and measured quantities. Entity alignment should therefore combine canonical well or project identifiers, spatial coordinates, geological context, and normalized units, rather than relying on textual similarity alone. For example, the same well appearing under an operator identifier and a research-project label should be merged only when location and supporting attributes agree. The LLM-Align approach developed by Chen et al. provides relevant general-purpose evidence for attribute- and relation-guided entity alignment, but its original evaluation is not geothermal-specific [33].

Cross-source ontology mapping should also specify allowed entity types and relations before integration. An alignment proposal can be treated as a candidate match A(e_i,e_j), whose acceptance requires an auditable set of identifier, location, and semantic checks. Where well names, formation classifications, or measurement conventions conflict, both candidates and their source records should be retained for expert adjudication. This step avoids silently converting a plausible semantic match into an asserted geological fact.

(3) Knowledge completion

Relation completion is distinct from verified knowledge acquisition. An LLM may suggest a missing relation from entity descriptions, graph neighborhoods, or retrieved reports, but a proposed edge should only enter the confirmed KG when its source span and ontology constraints are satisfied. Structure-aware and semantic-fusion approaches to KG completion have been investigated in general settings [34]. For geothermal applications, suggestions such as a fault intersecting a reservoir, or a monitoring event affecting a named well, require site-specific documentary or geospatial support; otherwise, they should remain explicitly flagged as hypotheses.

(4) KG–RAG collaborative reasoning and geothermal ontology example

A proposed geothermal KG–RAG workflow can be organized into four interacting operations. First, KG-enhanced retrieval resolves entities in a user query, follows permitted well–reservoir–measurement relations, and uses the resulting graph neighborhood to constrain passage retrieval. Second, RAG-assisted KG completion retrieves supporting source spans for candidate missing relations and proposes updates for expert or rule-based checking. Third, joint reasoning combines graph paths with retrieved technical passages to generate an answer with explicit citations, while preserving the distinction between a recorded fact and an inferred relation. Fourth, incremental updating re-runs entity alignment and conflict checks as new drilling or monitoring reports arrive. Related KG-RAG question decomposition and graph-grounded QA mechanisms have been evaluated in general-domain research [23], while Zhou and Li (2025) provide geoscience evidence for KG construction and incremental LightRAG updating [26].

A compact representation makes evidence handling explicit. Let an asserted fact be $f = (s, p, o, \tau, \pi, c)$, where s and o are the subject and object, p is an ontology-defined relation, τ is the valid time or observation interval, π identifies the source document and passage, and c is an evidence-support score. Each numerical measurement additionally records units, depth interval, and measurement method. Candidate facts should undergo entity alignment, schema checking, source verification, and unit normalization before incorporation. A conflict is flagged when two facts concern the same normalized entity and property over overlapping valid intervals but provide incompatible values beyond reported measurement uncertainty, or assert mutually exclusive relations. Conflicts must be preserved with both provenance records; preference for a newer document is appropriate only when observation time, document authority, and revision status establish that it supersedes the earlier statement.

Confidence propagation should distinguish evidence support from model self-confidence. A practical conservative rule is that the support assigned to a multi-hop reasoning path cannot exceed that of its least-supported factual link unless independently corroborated; scores from duplicate or dependent reports should not be treated as independent evidence. The score c should not be interpreted as a calibrated probability until calibrated on held-out, expert-labeled geothermal cases. When source support, temporal validity, or cross-document consistency is insufficient, the system should report uncertainty or abstain from generating a definitive engineering recommendation. Temporal graph retrieval has been studied in general temporal knowledge-graph QA [35], but its calibration and safety consequences in geothermal settings remain to be established.

For example, a minimal geothermal ontology may include the classes GeothermalWell, Reservoir, LithostratigraphicUnit, Fault, Measurement, OperationalEvent, and SourceDocument, with relations such as intersects, locatedIn, measuredAt, affects, and reportedIn. The question "Which wells intersect a fault and have temperatures exceeding a specified threshold during the latest survey?" would first resolve well and fault identifiers, traverse spatial relations, retrieve temperature measurements together with units and timestamps, and return only records with traceable sources. If a well-log temperature conflicts with a recent survey or uses a different depth interval, the discrepancy is exposed rather than collapsed into a single value. Evaluation of such a workflow should separately report entity-alignment precision, relation-extraction F1, conflict-detection recall, temporal validity, source attribution, and expert-adjudicated answer correctness.

3.1.4. Enhanced Data Infrastructure and Semantic Retrieval

Reliable geothermal AI depends on consistent metadata, source provenance, quality assessment, and interoperable domain datasets. Neumann et al. (2026) reported a comprehensive update to the Global Heat Flow Database, including standardized metadata, measurement quality controls, and traceable reporting of geothermal heat-flow observations [36]. Such infrastructures provide structured inputs for geothermal knowledge retrieval and subsequent AI integration.

Geological-record standardization workflows [28] and curated geothermal data catalogues with explicit metadata [36] provide complementary foundations for knowledge-graph construction, structured retrieval, and source-grounded explanations. Integration with RAG requires separate assessment of data coverage, measurement-unit consistency, temporal updating, and retrieval provenance.

3.2. Development Parameter Prediction and Decision Support for Optimization

3.2.1. Prediction of Key Geothermal Reservoir Parameters

Quantitative geothermal reservoir modeling depends on measured temperatures, permeability, and heat-flow constraints. Conventional machine learning has been used to estimate geothermal simulator input parameters: Suzuki et al. (2024) trained models to estimate permeability and heat source or sink conditions using simulator-generated data, and assessed the resulting temperature distributions [37]. Separately, Al-Fakih et al. (2025) used Bayesian-optimized regression models to estimate geothermal temperatures in western Yemen [38]. Both studies evaluate conventional machine-learning workflows rather than LLM-augmented models; their reported predictive metrics therefore do not demonstrate that LLM-derived semantic features improve geothermal reservoir-parameter prediction.

(1) Geothermal reservoir property modeling through the integration of textual semantic information

LLMs could be used to extract structured descriptions of lithology, fracture occurrence, or alteration from geological documents, drawing on scientific information-extraction methods demonstrated in adjacent domains [30,39]. A hybrid workflow would then match these descriptions to measured borehole or field data before a specialized physical or statistical model performs quantitative estimation.

Geological narratives describing lithology, fractures, and hydrothermal alteration could be aligned with measured reservoir properties by well, depth, and observation date. Text-derived fracture indicators do not have universal positive relationships with permeability, thermal conductivity, or temperature anomalies. Permeability depends on fracture aperture, connectivity, infilling, and stress conditions. Rock thermal conductivity is influenced by mineral composition, porosity, fluid saturation, temperature, and rock structure, as supported by geothermal-rock experiments [40]. Temperature anomalies further reflect heat supply, fluid flow, and local boundary conditions. Figure 5 outlines a conceptual workflow integrating text-derived geological information, measured reservoir properties, physical constraints, and independent validation.

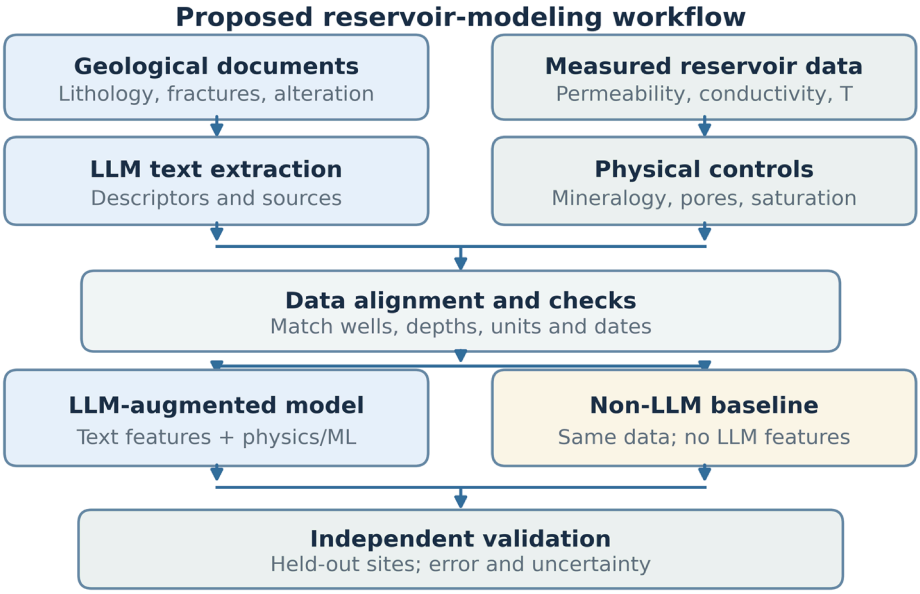


Figure 5. Conceptual workflow for integrating LLM-derived geological descriptions with measured reservoir data, physical constraints, and independent validation.

Recent reviews of machine learning for subsurface geothermal applications identify opportunities for data-driven reservoir characterization alongside challenges related to geological complexity and validation [41]. The added predictive value of LLM-derived textual indicators could be assessed using paired geological descriptions and measured reservoir properties, appropriate non-LLM baselines, uncertainty estimates, and independent geoscientific validation.

Uncertainty-aware reservoir inversion should separate the uncertainty of inferred parameters from uncertainty in the predicted reservoir response. For estimated permeability, temperature, or fracture properties, Bayesian inversion may report parameter posterior credible intervals, while frequentist procedures may produce confidence intervals under an explicitly defined sampling model. Forward simulations or validated surrogates then propagate those parameter distributions into predictive intervals for temperature drawdown, production performance, or pressure evolution. The geothermal history-matching study of Ullah et al. (2024) illustrates uncertainty-oriented inverse modeling with a Bayesian particle-swarm/BiGRU workflow, but it is not an LLM validation study [3]. An LLM can retrieve observations, organize assumptions and communicate these intervals, yet cannot

substitute its self-reported confidence for calibrated parameter or output uncertainty. Independent well- or site-level testing, interval coverage and physical consistency should be checked before using the outputs in reservoir design.

(2) LLM-Based Rapid Resource Potential Screening and Issue Identification

Beyond reservoir-scale prediction, LLMs can assist in knowledge integration and exploratory research analysis. Hatvani et al. (2024) elicited priority research questions in Earth and geosciences using LLM prompts and identified sustainable geothermal energy development among recurring themes [42]. Changes between closely related prompt terms altered the generated questions, illustrating sensitivity in thematic elicitation.

3.2.2. Drilling Engineering Optimization and Risk Early Warning

Drilling is one of the most expensive and risk-intensive stages of geothermal development. LLMs may assist with the interpretation of drilling records and operational alerts; however, evidence from petroleum wells or generic language-processing benchmarks cannot establish the reliability of real-time geothermal optimization or early warning. Differences in high-temperature conditions, fractured loss zones, reporting practices, and operating constraints must be considered before such systems are tested in the field.

(1) Automated identification of drilling dysfunctions

Drilling-risk methods differ by input data and model class. Conventional natural language processing (NLP) methods extract or classify information from narrative drilling reports, whereas conventional machine-learning (ML) methods forecast abnormal events from structured sensor and drilling-telemetry time series. Gurina et al. (2022) used telemetry-based conventional ML to forecast abnormal drilling events in oil and gas wells [43].

LLMs could potentially connect contextual information in unstructured drilling descriptions, such as a rise in torque accompanied by changes in pump pressure. Their contribution to extraction or warning accuracy should be assessed against conventional NLP and ML baselines using independently labeled geothermal records, particularly under variable lithology, high-temperature conditions, different sensor configurations, and rare hazard events. Safety-critical warning performance should report missed hazards and false alarms separately rather than relying on overall classification accuracy.

(2) Real-time parameter optimization and risk early warning

Conventional drilling-risk studies have used drilling measurements, including weight on bit (WOB), rotational speed (RPM), torque, and drag, to evaluate operating conditions and abnormal events [43,44]. An LLM-assisted workflow could link such telemetry to engineering documents and case histories, while leaving quantitative drilling optimization to validated analytical or control models. Such an interface would need to accommodate incomplete or delayed measurements, location-specific loss-circulation and wellbore-instability mechanisms, and constraints governing drilling fluids, tools, and allowable operating ranges.

An LLM could organize evidence, flag uncertain operating states, and present candidate actions for review by drilling personnel. Recommendations concerning drilling-fluid changes, bit selection, or operating adjustments would require independent checks against site-specific physical models, verified equipment limits, and approved procedures.

For drilling risk warning, a practical uncertainty-aware output would couple an LLM-extracted event description and source evidence with a separately validated, calibrated probability of the relevant hazard. The risk model should be evaluated on held-out geothermal wells using the precision–recall curve, event-level recall, missed-hazard rate, Brier score or reliability diagrams, and warning lead time. Thresholds should reflect the relative consequences of missed hazards and false alarms, site-specific response times, and

prescribed escalation procedures, rather than LLM self-confidence or values transferred from petroleum datasets. Low-quality or missing sensor data should trigger abstention or expert review.

3.3. AI-Assisted Engineering Design and Dynamic Optimization

3.3.1. Automated Numerical Simulation and Code Generation

Numerical simulation is a fundamental tool for geothermal reservoir analysis and engineering design. However, developing, optimizing, and maintaining simulation codes typically requires extensive programming expertise and considerable development time. Recent advances in LLMs have demonstrated substantial potential for improving the automation and efficiency of geothermal numerical simulation workflows.

(1) LLM-assisted parallelization of geothermal simulation codes

Peer-reviewed work on LLM-driven physics simulations demonstrates the feasibility of translating natural-language tasks into simulator interactions and using iterative feedback to correct intermediate results [45]. However, this evidence comes from physics simulation outside geothermal reservoir engineering. An application involving geothermal numerical codes would still need to verify solver interfaces, reproducibility, unit conversions, numerical convergence, and thermal or hydraulic conservation independently.

(2) Automated workflow construction and script generation

LLMs could assist in drafting scripts for data preprocessing, result visualization and interfacing with specialized geothermal simulators. In a peer-reviewed adjacent-domain physics-simulation study, Park et al. (2026) examined a self-correcting LLM workflow with simulator feedback [45]. This does not show that a generated TOUGH2 or other geothermal solver workflow is correct. Any candidate code should undergo executable tests, benchmark-case comparisons, boundary-condition checks and physical-conservation verification before engineering use.

3.3.2. Engineering Design Assistance and Conceptual Solution Development

During the conceptual design stage, LLMs can serve as both creative ideation tools and knowledge integration assistants, helping engineers explore alternative design strategies and synthesize multidisciplinary information during the early phases of project planning.

(1) LLM-assisted conceptual design of large-scale geothermal well arrays

LLMs may assist at the conceptual stage by organizing candidate geothermal well layouts, identifying engineering constraints, and outlining design alternatives. Their role is then a structured knowledge interface rather than a validated reservoir optimizer. Well spacing, thermal interference, drilling costs, production rates and lifecycle economics must still be established by specialized models and independently checked evidence. The general possibilities of AI-supported engineering planning do not by themselves establish quantitative geothermal design performance.

(2) Comparative analysis of integrated multi-energy system configurations

The design of integrated geothermal energy systems involving heating, power generation, cooling, and other coupled energy services requires comprehensive evaluation of technological alternatives. Retrieval-based energy knowledge systems offer relevant information-synthesis methods [27], while research on AI-supported distributed energy planning provides complementary optimization context [46]. LLMs may help organize manuals, engineering cases, and policy documents during early-stage screening. However, performance established in broader integrated-energy applications does not directly demonstrate geothermal design validity: reservoir temperature and deliverability, reinjection requirements, local heat-demand profiles,

equipment operating envelopes, and jurisdiction-specific permitting must be incorporated into the evaluation. Accordingly, LLM-generated comparisons should be used to frame candidate configurations for subsequent techno-economic, thermodynamic, and expert assessment rather than as verified designs.

3.3.3. System Architecture Selection and Cost–Benefit Trade-Offs

When deploying LLM-based applications, architectural complexity should be evaluated against task performance and cost. Sabbagh et al. (2024) compared single-agent and multi-agent LLM systems for oil and gas well-construction questions and Text-to-SQL tasks [47]. Their journal study reported that the multi-agent configuration improved question-answering truthfulness by 28% relative to the single-agent configuration but incurred approximately 3.7 times the cost, whereas the single-agent configuration achieved a 15% higher score on Text-to-SQL tasks. These results concern different tasks and do not establish that a particular architecture is universally superior. Selection for geothermal applications should be based on independent geothermal benchmarks, cost constraints, and sensitivity to task-specific error types.

3.3.4. Multi-Source Information Fusion and Spatial Semantic Enhancement

Geothermal engineering decisions frequently require reasoning over geographic relationships and technical descriptions. Liang et al. (2026) proposed GeoAgentic-RAG, a multi-agent framework combining semantic–spatial retrieval, executable geographic analysis, and multimodal information [48]. The system was evaluated on geospatial reasoning tasks in Nanjing and Guangzhou. A potential geothermal adaptation would connect mapped faults, borehole locations, geophysical surveys, and protected land parcels, with spatial query logic, data quality, and engineering suitability independently verified.

Spatial–semantic retrieval is one component of a broader multimodal workflow. A geothermal foundation model would also require aligned representations of seismic amplitude volumes, depth-indexed well logs, core-image features, and time-dependent simulation fields, while retaining original measurement units, sample provenance, and spatial resolution. Unlike text-only or GIS-based RAG, a multimodal design should retrieve documentary passages together with corresponding image patches, log intervals, or field slices, using cross-modal evidence links and specialized decoders for non-text inputs. Section 4.2 outlines the corresponding integration architecture and validation requirements [24,49].

To summarize the characteristics of the above applications, Table 2 presents representative scenarios in which LLMs support geothermal engineering design, comparing them in terms of application stage, LLM functionality, input information, output results, and current limitations.

Table 2. Typical application scenarios of LLMs in geothermal engineering design.

Design Stage	Main Task	LLM Technique	Input Information	Output	Added Value	Key Limitations
Numerical simulation assistance	Model construction and code development	GPT/code LLMs	Governing equations, boundary conditions	TOUGH2/ FEFLOW/ Python code	Lowers modeling barriers and improves development efficiency	Requires validation of physical consistency
Site selection	Knowledge retrieval and constraint identification	Retrieval-Augmented Generation (RAG)	Geological data, regulations, case studies	Site evaluation indicator system	Enhances information acquisition efficiency	Limited capability to replace quantitative assessment

Table 2. *Cont.*

Design Stage	Main Task	LLM Technique	Input Information	Output	Added Value	Key Limitations
Conceptual design	Design scheme generation	Prompt engineering	Engineering objectives, boundary conditions	Well layout and operation schemes	Expands design space exploration	Limited optimization capability
Technology comparison	Multi-scheme analysis	RAG + Chain-of-Thought (CoT) reasoning	Technical standards, case databases	Comparative evaluation of technologies	Supports decision analysis	Highly dependent on knowledge base quality
Parameter optimization	Optimization recommendation generation	LLM + machine learning (ML)	Historical operational data	Parameter adjustment suggestions	Provides empirical decision support	Cannot replace numerical optimization algorithms
System deployment	Intelligent knowledge services	Agent-based systems + RAG	User queries	Intelligent QA systems	Lowers barrier to knowledge access	Susceptible to hallucination risks

3.4. Chapter Summary

This chapter reviewed LLM-related approaches for geothermal knowledge management, document processing, reservoir-analysis support and engineering planning. The strongest peer-reviewed application evidence in the studies examined here concerns particular tasks evaluated in geology, petroleum engineering and broader energy systems. Geological retrieval, information extraction and knowledge-graph reasoning offer potentially useful components for geothermal document workflows, but require testing against geothermal terminology, records and expert-reviewed reference answers. Studies of conventional machine learning for reservoir prediction and of numerical simulation should not be counted as empirical validation of LLM-enabled geothermal parameter estimation or risk warning.

Across the reviewed studies, an important distinction emerges between the demonstrated functions of individual LLM components and the performance of a complete geothermal decision-support system. RAG may support traceable access to documents; cross-modal integration and tool calling offer promising engineering interfaces; and physical simulators remain responsible for quantitative predictions and conservation constraints. However, reported scores are task- and dataset-specific and cannot establish field readiness without independent geothermal evaluation. Section 4 therefore concentrates on standardized benchmarking, transfer boundaries, uncertainty-aware decisions and methods for combining LLMs with verified engineering models.

4. Discussion

4.1. Challenges and Limitations

Large language models have been applied to geothermal knowledge services and investigated in adjacent engineering and geoscience domains, whereas robust evidence for their use in high-stakes geothermal decisions remains limited. Table 3 summarizes the evaluated tasks, reported evidence, key limitations, and geothermal validation needs of six selected peer-reviewed journal studies. All six were tested in adjacent domains rather than on independently validated geothermal tasks. Table 4 presents the corresponding published quantitative results in their original task settings. Because the datasets, metrics, and evaluation procedures differ, these studies cannot be compared through a common performance ranking or used to infer equivalent geothermal engineering reliability.

Table 3. Evidence and validation characteristics of representative LLM applications relevant to geothermal engineering.

Application Scenario	Representative Study	LLM-Based Approach	Evidence Type	Reported Evaluation Evidence	Key Limitations	Geothermal Validation Requirements
Historical well-record extraction	Ma et al. (2024) [32]	Llama 2 extraction of location and depth from well records	B	160 orphaned-well documents; reported 100% end-to-end workflow accuracy for clean PDFs and 70% for image-based records.	Oil/gas records; sensitivity to input document quality.	Validate on independent geothermal drilling logs, including scanned records and error analysis.
Energy-domain question answering	Chebbi and Kolade (2026) [25]	EnergyGPT: LLaMA 3.1-8B with supervised and LoRA adaptation	B	Energy-domain QA benchmarks and comparisons with the base model; human and LLM-judge assessments.	Not a geothermal-specific evaluation; possible domain and assessment effects.	Use expert-annotated geothermal QA sets, source-grounding checks, and independent held-out tests.
Geoscience KG-assisted QA	Zhou and Li (2025) [26]	Geoscience LLM with knowledge-graph construction and LightRAG	B	Mineral-prospecting entity extraction and geological QA experiments; model-based answer assessment.	Mineral-exploration scope; potential bias from an LLM evaluator.	Verify geothermal entities and relations, retrieval evidence, and expert-rated answer correctness.
Well-engineering LLM architecture	Sabbagh et al. (2024) [47]	Single- and multi-agent architectures for well-construction QA and Text-to-SQL	B	In oil/gas tasks, multi-agent QA was more truthful (+28%) but costlier (3.7-fold); single-agent Text-to-SQL scored higher (+15%).	Performance and costs depend on the task and the petroleum-domain corpus.	Repeat task-controlled evaluation on geothermal well records and measure accuracy, error and cost.
Geological data standardization	Yan et al. (2026) [28]	RAG-based LLM for formation-type standardization of borehole descriptions	B	Expert-reviewed data: 100 boreholes from 21 Macao projects (908 samples); evaluated standardization and interpretability.	Urban geotechnical context; potential dependence among boreholes from the same projects.	Test site-/project-separated geothermal boreholes, missing modalities and formation-specific errors.
Domain-specific geological reasoning	Ma et al. (2026) [29]	QHGeoGPT: LoRA tuning, RAG and geological knowledge-graph reasoning	B	1201 expert-designed questions in mineral exploration; reported accuracy of 89.99%.	Mineralization-specific tasks and ontology; no geothermal-specific engineering validation.	Test on geothermal reservoir terminology and reasoning, with independent expert checks.

Note: B = empirical validation in an adjacent domain, not in a geothermal-specific task. Reported evaluation results concern different tasks and datasets and should not be interpreted as a common performance ranking.

Table 4. Reported quantitative evidence from selected journal studies relevant to geothermal LLM applications.

Study/Domain	LLM/Method	Reported Data Scale	Evaluation Metrics	Reported Findings
Ma et al. (2024) [32]/Oil and gas	Llama 2 information extraction	160 historical orphaned-well records	Field extraction accuracy by document format	Complete document-conversion-plus-LLM workflow: 100% for clean PDFs vs. 70% for unstructured image-based records.
Chebbi & Kolade (2026) [25]/General energy	EnergyGPT: LLaMA 3.1-8B, SFT vs. LoRA	476 energy questions: 233 MCQ, 100 T/F, 143 open-ended	MCQ and T/F accuracy; human evaluation of open answers	MCQ: base 73%, SFT 86%, LoRA 89%; T/F: base 95%, SFT and LoRA both 94%.
Zhou & Li (2025) [26]/Mineral geoscience	GeoGPT + LightRAG and KG construction	Geological entity extraction and mineral-domain QA	Mean entity-extraction F1; model-judged QA win rates	Mean entity F1 = 0.8353 (vs. Qwen2.5 0.785 and DeepSeek-V3 0.729 under domain prompts).
Sabbagh et al. (2024) [47]/Oil and gas	Single-agent vs. multi-agent LLM architectures	Well-construction QA and Text-to-SQL tasks	QA truthfulness; task score; relative cost	Multi-agent QA truthfulness +28% at about 3.7x cost; single-agent Text-to-SQL score +15%.
Yan et al. (2026) [28]/Engineering geology	Stratified RAG for formation-type standardization	100 boreholes; 21 projects; 908 formation samples	Formation classification; input-strategy comparison; token cost	For the Macao test data, the all-to-individual strategy was reported as more robust and token-efficient; no geothermal benchmark.
Ma et al. (2026) [29]/Mineral geoscience	QHGeoGPT: LoRA + RAG + KG reasoning	1201 expert-designed geological QA questions	Domain QA accuracy; terminology coverage	89.99% QA accuracy; 62.79% terminology coverage in mineral-exploration tasks.

Note: Values and performance metrics are reproduced for the tasks tested in the original journal studies. All six studies were conducted in adjacent petroleum, energy, geoscience, or geotechnical domains (Evidence Type B), not in independent geothermal validation. MCQ = multiple choice; T/F = true/false; F1 = harmonic mean of precision and recall; SFT = supervised fine-tuning; KG = knowledge graph. Different task metrics, samples, and evaluation procedures are not directly comparable.

The evidence in Table 3 is strongest for the existence of task-specific evaluation in adjacent domains—not for validated transfer to geothermal operations. Reported tests cover well-record extraction, energy-domain question answering, knowledge-graph-assisted geological reasoning, geological data standardization, and architectural trade-offs in well engineering. Their differences in datasets, evaluators, and evaluation criteria preclude a direct ranking of the methods. Geothermal-specific held-out test sets, independent expert verification, documentation of training–test separation, and engineering consequence analysis are therefore required before performance claims can be extended to geothermal applications. The quantitative findings in Table 4 should be interpreted within each source study rather than as a unified geothermal benchmark.

To improve comparability across these heterogeneous case studies, a task-oriented geothermal LLM benchmark framework is proposed. Drawing on principles of multi-dimensional and transparent language-model evaluation [50,51], the framework covers knowledge question answering, information extraction, LLM-assisted reservoir parameter prediction, drilling-risk identification, and numerical-model code generation. Table 5 specifies candidate data inputs, reference annotation requirements, and task-specific metrics.

Table 5. Proposed task-specific benchmark framework for geothermal LLM applications.

Task	Candidate Data Sources and Inputs	Reference Annotation Standard	Suggested Evaluation Metrics	Critical Validation Checks
Knowledge Q&A	Public GDR reports, geothermal journal articles, wellfield documentation, and accessible technical standards.	Expert-approved answer; passage-level supporting evidence; acceptable alternatives and explicit unanswerable cases.	Answer correctness or expert-scored F1; citation precision; evidence faithfulness; abstention accuracy.	Test out-of-scope queries, conflicting sources and citation support; distinguish RAG retrieval from generation.
Information extraction	Geological assessments, lithology narratives, drilling/completion reports and well-log interpretations.	Entity spans and typed fields (formation, depth, temperature, event); normalized units; source coordinates.	Entity/relationship precision, recall and F1; numerical-field error; unit-consistency rate.	Include scanned/OCR reports; audit near-duplicate logs, missing values and provenance.
Reservoir parameter prediction	Site-disjoint well measurements, reservoir assessments and documented simulation scenarios.	Measured or independently verified tempera-ture/permeability targets, units, locations, uncertainty and solver settings.	MAE, RMSE, R^2 and interval coverage where applicable; physical-constraint violation rate.	Evaluate the full LLM-assisted predictor, not LLM text alone; held-out site tests and physics-based checks.
Drilling risk identification	Time-stamped geothermal drilling logs, incident reports and associated operating measurements.	Event type, onset interval, severity, evidence excerpt and expert-reviewed no-event labels.	Precision/recall, PR-AUC, critical-event miss rate, false-alarm rate and detection lead time.	Cost-sensitive thresholds; delayed labels; independent review of high-consequence false negatives.
Code generation	Public geothermal simulator examples, solver input files, governing equations and unit-tested scripts.	Executable reference tests, numerical tolerances, expected outputs, units and conservation constraints.	Test-pass rate, execution success, numerical deviation, conservation residual and computational cost.	Isolated execution; reproducible solver versions; expert verification before engineering reuse.

Note: GDR = Geothermal Data Repository; Q&A = question answering; MAE = mean absolute error; RMSE = root mean square error; PR-AUC = area under the precision–recall curve. Numerical evaluation targets require independently measured or physically verified reference values.

Benchmark construction should combine permission-cleared geothermal data with traceable metadata and a fixed annotation manual. For each task, the manual should define domain terminology, entity and relationship types, physical units, allowable numerical tolerances, citation spans, ambiguous cases, and an explicit "insufficient evidence" label. Independent geothermal specialists should annotate representative benchmark items, with documented disagreement resolution and adjudication. Project-disjoint and site-disjoint train/development/test splits, temporal holdouts, and checks for duplicated or potentially pretrained materials would help expose leakage and domain-shift risks.

Model comparisons should use identical versioned splits, document prompts, retrieval corpora, tool access, solver versions, cost and latency, and separately report accuracy, source attribution, abstention and failure modes. For high-risk tasks, engineering specialists should independently inspect stratified errors and safety-critical outputs using a predefined rubric; missed drilling hazards and violations of physical constraints should be reported separately from aggregate scores. As an implementation pathway, future community-maintained leaderboards could publish dataset cards, evaluation scripts and baseline results while

reserving controlled-access hidden tests for final ranking. Results should be reported by task, reservoir setting and validation level, without implying that offline benchmark performance constitutes field deployment evidence.

From a technical perspective, current geothermal applications of LLMs face three progressively interconnected challenges: limited domain data, model reliability, and multimodal information integration. The first and most fundamental limitation is the scarcity of high-quality domain-specific data. Publicly available geothermal datasets remain limited in both quantity and coverage, while existing data are distributed across multiple sources and stored in heterogeneous formats. A substantial proportion of critical domain knowledge is still embedded within unstructured exploration reports, drilling records, and engineering documents. Moreover, many high-value operational datasets are restricted by commercial confidentiality, forcing most studies to rely on relatively small datasets for model development and validation. Because geothermal systems exhibit strong regional heterogeneity, models trained on limited local data are prone to learning site-specific patterns, resulting in insufficient cross-regional generalization.

The second major challenge concerns model reliability and hallucination, which constitutes one of the principal obstacles to engineering deployment. Since LLMs generate outputs based on statistical correlations rather than explicit physical principles, they may produce factual inaccuracies, fabricated engineering parameters, or logically inconsistent conclusions [52]. In high-risk applications, such as geothermal resource assessment and engineering design, these errors may directly influence critical engineering decisions. Although approaches such as Retrieval-Augmented Generation (RAG) and Knowledge Graph (KG) constraints can substantially reduce the occurrence of hallucinations, they cannot fundamentally guarantee the physical consistency or engineering validity of generated results.

A related limitation is the inadequate quantification of task-specific uncertainty. Aleatoric effects from reservoir heterogeneity and measurement noise differ from epistemic effects associated with incomplete data and uncertain models [53,54]. A language model's verbal confidence is neither a calibrated hazard probability nor an interval estimate of reservoir parameters. Section 4.3 therefore distinguishes uncertainty methods by the quantity they estimate and the engineering decisions they can support.

Furthermore, multimodal geothermal data remain difficult to align and validate. Seismic volumes, well-log curves, core images, numerical fields, and engineering reports differ in spatial scale, depth reference, timestamps, file formats, and measurement uncertainty. Paired annotations linking observations from the same interval or event are often sparse, while missing modalities and asynchronous monitoring further complicate fusion. These challenges motivate the four-layer conceptual roadmap in Section 4.2 (Figure 6) [24,55].

Overall, the selected journal evidence supports specific LLM-enabled methods in adjacent fields more strongly than it supports verified deployment under geothermal engineering conditions. Domain shift, task-specific uncertainty, technical terminology, and site-level physical and regulatory requirements limit direct transfer. Strengthening the evidence base therefore requires transparent benchmarks, structured assessment of transfer applicability (Section 4.2), and staged validation with specialist oversight.

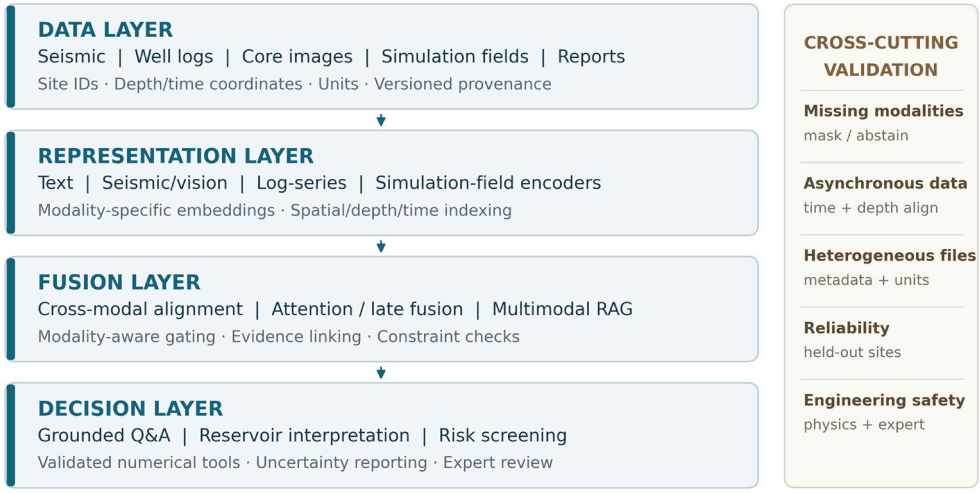


Figure 6. Proposed four-layer technical roadmap for geothermal multimodal foundation models.

4.2. Technical Integration and Engineering Applicability

Cross-domain transfer applicability assessment. Petroleum engineering, broader geoscience, and energy-system studies can inform geothermal LLM design, but source-domain performance does not establish geothermal validity. A transfer assessment should distinguish four complementary dimensions: data distribution, terminology systems, physical processes, and engineering constraints. The broader problem of performance under distribution shifts is well documented in domain-generalization research, while geothermal drilling reviews identify operating conditions that depart substantially from conventional oil-and-gas applications [56,57].

Data distribution. Petroleum and geothermal datasets differ in lithology, temperature and pressure distributions, fracture characteristics, document formats, label prevalence, and measurement or reporting frequency. A model trained on petroleum well-construction reports may learn correlations that do not hold for geothermal production and reinjection wells. Before adaptation, source and target corpora should be profiled for missing fields, class imbalance, temporal coverage, and geographic or operator effects. Evaluation splits should hold out entire geothermal wells, projects, and sites to limit leakage between related records.

Terminology systems. Similar-looking descriptions may denote different operational phenomena, and common terms may refer to different thresholds or procedures across drilling practices. Transfer therefore requires a geothermal terminology and event ontology that maps lithological units, fracture and feed zones, lost-circulation events, drilling-fluid properties, well configurations, and measurement units. Expert-labeled geothermal samples should test entity matching, relation extraction, unit normalization, and consistency with project records. Ambiguous expressions should trigger source-document retrieval or abstention rather than unconstrained inference.

Physical processes. High-temperature geothermal wells and engineered reservoirs involve thermal stresses, fractured rock flow, fluid–rock reactions, and coupled thermal–hydraulic–mechanical–chemical processes that may be absent or underrepresented in source-domain data. The behavior of lost-circulation materials, for example, is temperature dependent under geothermal conditions [58]. Transferred recommendations concerning well integrity, pressure management, or reservoir responses must therefore be checked against the appropriate governing equations, boundary conditions, materials, and numerical or laboratory evidence; semantic plausibility alone does not establish physical validity.

Engineering constraints. Geothermal applications must respect well-specific operating limits, high-temperature equipment specifications, fluid chemistry and corrosion risks, reinjection and induced-seismicity controls where applicable, and jurisdiction-specific environmental, permitting, and safety requirements. A model that retrieves a technically plausible petroleum drilling procedure may still suggest an action inconsistent with geothermal site conditions or the applicable standards. Standards-aware retrieval with document-version tracking, explicit constraint checks, and approval by qualified engineers are necessary safeguards before operational use [57,59].

Suggested transfer-validation protocol. A reproducible study could assemble permission-cleared petroleum source data and an independently annotated geothermal target corpus spanning different reservoirs, drilling contractors, record formats, and temperature regimes. With identical held-out geothermal test projects, investigators should compare (i) an unadapted source-domain model, (ii) geothermal terminology and retrieval augmentation, and (iii) parameter-efficient adaptation using a separate geothermal training partition, alongside appropriate specialist or conventional-model baselines. Blinded experts should adjudicate ambiguous annotations and high-consequence errors. Evaluation should report entity-level precision, recall and F1, numeric/unit exact-match rates, hallucinated constraints, calibrated risk scores, missed-hazard and false-alarm rates, and performance by site and document type, with confidence intervals obtained through project-level resampling.

Before field use, candidate systems should also undergo scenario-based stress tests for unfamiliar formations, incomplete measurements, loss-circulation events, and conflicting or outdated regulatory documents. Physical consistency and compliance checks should be evaluated independently of linguistic fluency, with pre-specified acceptance rules for safety-critical errors and mandatory human review of uncertain outputs. Failure to meet target-domain criteria should restrict an LLM to traceable document retrieval and explanation instead of autonomous engineering recommendations.

The same transfer safeguards apply to KG–RAG collaborative reasoning (Section 3.1.3). Entity alignment, relation completion, source-aware conflict detection, and temporally valid graph retrieval should be evaluated on geothermal well records and technical standards rather than inferred from mineral-exploration or generic QA results. Any confidence attached to graph paths must be calibrated using independent domain annotations, and unresolved discrepancies must trigger expert review before a system offers safety-relevant advice.

Technical roadmap for geothermal multimodal foundation models. Figure 6 synthesizes a four-layer architecture comprising a data layer, representation layer, fusion layer and decision layer. The data layer assembles permission-cleared seismic data, depth-indexed well logs, core images, simulation fields, and geological or operational text. Each observation requires a common well/site identifier and explicit spatial coordinates, depth datum, acquisition time, measurement units, quality flags and provenance. Rather than forcing all data into an unstructured prompt, the representation layer applies modality-specific encoders (for example, seismic and core-image vision encoders, sequence encoders for logs, field encoders for simulated states, and text encoders for reports) and preserves references to original files. Recent multimodal seismic interpretation research provides an adjacent-geoscience example of combining seismic representations with well-log prompts, but it does not establish a geothermal-specific LLM solution [24].

Cross-modal alignment and fusion. Matching observations should be constructed using well identifiers, surveyed coordinates, stratigraphic/depth intervals and acquisition times, with geophysical registration and explicit resolution conversion. Paired samples can support contrastive alignment or cross-attention across modality embeddings; late fusion with quality-weighted gates may be preferable when pairing is sparse. The fusion layer

would combine these representations with multimodal retrieval-augmented generation (RAG), retrieving linked text passages, log intervals, image patches and simulation outputs while preserving evidence provenance. Parameter-efficient vision–language fine-tuning could be considered only when task-specific aligned examples and independent evaluation are available. Graph or spatial relationships may act as constraints or retrieval indices, but do not themselves guarantee physical correctness.

Missing, asynchronous and heterogeneous inputs. A field-ready design must tolerate absent core photographs, incomplete logging curves, sensor outages and delayed simulation updates. Modality-presence masks, modality dropout during training, quality-aware fusion and fallback to the available evidence can limit dependence on complete inputs; any imputed modality should be explicitly flagged as inferred, never presented as an observation [55]. Streaming data should be time-stamped, aligned to defensible acquisition windows and depth references, and handled with explicit latency or stale-data flags. Heterogeneous formats (e.g., seismic volumes, well-log files, images, meshes and documents) require typed metadata, unit conversion and version-controlled transformation records rather than direct format-blind concatenation.

Decision layer and validation. The decision layer should provide provenance-linked geological question answering, reservoir interpretation support, anomaly screening, and candidate engineering actions, with quantitative reservoir responses delegated to verified numerical solvers. The proposed roadmap should be evaluated on site-disjoint geothermal cases with expert-linked cross-modal annotations, testing alignment accuracy, retrieval relevance, missing-modality robustness, temporal freshness, calibration of reported uncertainty, and physical/engineering constraint violations. Ablation studies should compare text-only retrieval, unimodal encoders, fused representations, and multimodal RAG under identical data splits. Uncertain or contradictory evidence should trigger abstention and expert review.

4.3. Uncertainty, Engineering Risks, and Future Research Directions

Future progress in geothermal LLM applications depends on uncertainty-aware engineering decisions and auditable integration with specialized physical models, rather than model scale alone. The following discussion distinguishes quantitative uncertainty in extracted information and model predictions from alternative coupling positions in physics-informed and simulator-assisted workflows.

Uncertainty quantification (UQ) must be tied to the specific output of an LLM-assisted geothermal workflow rather than treated as a generic confidence score. The main targets are (i) the correctness and source support of extracted facts and retrieved documents, (ii) uncertainty in reservoir parameters inferred from sparse geological and monitoring data, (iii) predictive uncertainty in thermo-hydraulic reservoir responses, and (iv) calibrated probabilities of drilling hazards. Bayesian inference, ensembles, and interval-producing regressors depend on different modeling assumptions. Conformal prediction provides distribution-free coverage under suitable calibration conditions [60], whereas probabilistic calibration assesses the reliability of predicted event probabilities [61]. Table 6 summarizes suggested methods and their application constraints.

Table 6. Uncertainty quantification methods and decision-support functions in geothermal LLM-assisted workflows.

Method	Uncertainty Output	Geothermal Task or LLM Role	Evaluation Criteria	Main Limitations
Bayesian inversion/ posterior sampling	Posterior credible intervals for reservoir parameters	Permeability, temperature, and fracture-parameter inversion supported by simulator data	Posterior predictive checks, coverage, sensitivity and identifiability	Prior choice, computational cost, model misspecification
Bootstrap and deep ensembles	Between-fit spread; empirical prediction intervals	Surrogate reservoir temperature or pressure forecasts; ensemble disagreement	Independent-well coverage, interval width, RMSE and out-of-distribution tests	Ensemble spread alone is not calibrated probability
Quantile/ heteroscedastic regression	Conditional prediction quantiles or estimated noise scale	Site-specific temperature or production-response prediction	Pinball loss, empirical interval coverage and width	Calibration can deteriorate under domain shift
Conformal prediction	Prediction intervals or sets with target marginal coverage	Calibrated intervals for validated numeric predictors and classified risk events	Observed coverage by site, horizon and risk subgroup; set width	Exchangeability assumptions; marginal not guaranteed conditional coverage
Monte Carlo uncertainty propagation	Distribution of simulated production and thermal responses	Propagating parameter and boundary uncertainty through THMC solvers	Convergence, sensitivity, scenario coverage, physical checks	Reliant on input distributions and model fidelity
Probability calibration for risk classification	Calibrated event probability, not LLM self-confidence	Lost circulation, wellbore instability and other drilling alerts	Brier score, reliability diagrams, PR-AUC, missed-event rate	Rare hazards, label quality and shifting operating conditions

Note: UQ methods differ in their target quantities and assumptions. A confidence interval for an estimated population quantity, a Bayesian credible interval for a model parameter, and a prediction interval for a future outcome have different interpretations. Conformal coverage generally assumes suitable exchangeability/calibration conditions [60].

Reservoir-parameter inversion. An operational study should state the inversion data, parameter bounds and prior assumptions, observation-error model, physical solver, and the target uncertainty quantity. A credible interval (Bayesian) or confidence interval (frequentist) for an inferred reservoir parameter must not be presented as a prediction interval for future production temperature. To inform a decision, uncertainty in the inferred permeability or heat-transfer properties should be propagated to simulated production-temperature and pressure trajectories, with results checked against withheld wells or subsequent monitoring records [3,6]. An illustrative decision rule is to flag a design for further simulation when the estimated probability of crossing a specified production-temperature limit exceeds an agreed risk tolerance. Neither the temperature limit nor the tolerance can be universal: they depend on project requirements, validated prediction models, and engineering review.

Drilling risk warning. For a hazard classifier, a calibrated event probability p may support an alert threshold selected using the costs of false alerts and missed hazardous events. Under a simplified two-action setting, with false-alarm cost C_FP and missed-

event cost C_FN and no other costs, an alarm is preferred when p exceeds C_FP/(C_FP + C_FN); actual drilling decisions also depend on hazard severity, lead time, baseline risk, intervention feasibility, and site regulations. Calibration should be checked on held-out geothermal wells and stratified by drilling phase and event type, using reliability diagrams and proper scoring rules [61]. Upper risk bounds or uncertain classifications may justify precautionary expert review; only independently validated models and approved protocols should be allowed to recommend operational actions.

Validation and safeguards. Bayesian or ensemble intervals require posterior/predictive checking and coverage assessment; conformal intervals provide marginal coverage only under their calibration assumptions and do not automatically remain valid under site-to-site or temporal distribution shift [60]. A complete workflow should test geologic outliers, incomplete log records, delayed measurements, contradictory source documents, and physical-conservation violations. An LLM should identify the underlying evidence and explain model-generated bounds without presenting token probabilities or fluent self-assessments as calibrated engineering uncertainty. Decision thresholds and escalation levels require independent validation through controlled geothermal cases and expert-reviewed trials.

LLM–physical model coupling paradigms. LLMs and physical models operate at different computational levels. An LLM can specify modeling objectives, translate technical text into solver inputs, select or call modeling tools, interpret intermediate results, and recommend candidate configurations; numerical solvers, physics-informed neural networks (PINNs), and neural operators perform the quantitative approximation or simulation. In particular, Deep Operator Network (DeepONet) [62] is a neural operator, not a language model. Existing geothermal PINN studies have modeled temperature, pressure and permeability under physical constraints [63], while a geothermal physics-constrained neural operator has provided accelerated performance evaluation [64]. Neither study alone demonstrates an integrated LLM–physical-model system. An LLM-driven iterative simulation framework has also been demonstrated in general multiphysics tasks, but geothermal-specific transfer remains untested [45]. Foundational methods for physically constrained learning include PINNs for forward and inverse PDE problems [65], physics-informed machine learning [66], and integration of scientific knowledge with predictive models [67].

Table 7 summarizes five potential LLM–physical model coupling architectures according to LLM positioning, information flow, interpretability, computational demand, and validation requirements. Reproducible interfaces should retain governing equations, initial and boundary conditions, mesh specifications, parameter units, solver versions, and simulation logs. Numerical results require independent verification beyond LLM-generated explanations.

Table 7. Classification of LLM–physical model coupling paradigms for geothermal reservoir simulation.

Coupling Paradigm/ Geothermal task	Coupling Position	Information Flow	Interpretability	Computational Cost	Validation Method
Input-oriented LLM + numerical solver Reservoir scenario setup	LLM front-end; THMC solver back-end	Objective/documents → checked input deck → numerical fields → summary	High if assumptions and inputs logged	Low LLM overhead; full solver runs costly	Unit, initial/boundary condition audits; solver reproduction

Table 7. *Cont.*

Coupling Paradigm/ Geothermal task	Coupling Position	Information Flow	Interpretability	Computational Cost	Validation Method
LLM + PINN Thermal/permeability inversion	LLM interface; trained PINN solves inverse problem	Well evidence → structured constraints → PINN fields/PDE residuals → explanation	Moderate; residuals and losses inspectable	PINN training costly; inference faster	Withheld wells; PDE/boundary residuals; simulator comparison [63]
LLM + neural operator Injection–production screening	LLM orchestrates pretrained operator surrogate	Controls + geology → operator response → LLM screening → solver check	Moderate; fields traceable, model internal opaque	High offline training; low inference	Out-of-distribution sites; error and physical-conservation tests [64]
Simulator-in-the-loop LLM Iterative reservoir design	LLM calls solver and consumes diagnostic feedback	Candidate → run → residuals/warnings → checked revision	High auditability with full logs	Multiple high-fidelity solver calls	Convergence; mass/energy balance; independent case replication [45]
Safety-gated multi-agent hybrid Planning and operational review	LLM agents around validated tools and human sign-off	Planner → solver/PINN → critic → domain expert → approval	Variable; evidence and roles must be recorded	High orchestration and verification overhead	Held-out field cases; agent ablations; expert/safety checks

Note: The coupling modes represent conceptual classifications. PINNs and neural operators are quantitative physical surrogates distinct from LLMs. Computational cost is qualitative and depends on training, solver fidelity, and number of runs. Arrows indicate the direction and sequence of information flow between components.

Illustrative geothermal reservoir simulation workflow. Consider a production–reinjection configuration with well-rate limits, estimated permeability structure, thermal properties, and monitored pressure/temperature data. An LLM could transform a natural-language question into a version-controlled reservoir model specification, while a verified TOUGH2-class simulator performs a high-fidelity baseline calculation. A PINN can provide an independently checked parameter-inversion result from monitoring and conservation constraints [63], and a pretrained neural operator can rapidly screen alternative well controls inside its validated parameter domain [64]. A simulator-in-the-loop controller could propose a candidate rate change, rerun the simulator, evaluate convergence and mass/energy balance, and either revise the candidate or return it to a geothermal specialist. In this proposed workflow the LLM manages information and tool orchestration; quantitative predictions and engineering acceptance remain subject to calibrated physical models and human authorization.

Coupling-specific validation. Front-end systems should be judged by correct extraction of parameter values, units, and boundary conditions; PINN pathways by held-out observations and physics-residual checks; neural operators by fidelity under new geological scenarios and against high-fidelity solutions; and iterative solver agents by reproducible trajectories, failure handling, conservation checks, and total compute expenditure. A practical comparison should measure the benefit against a manually configured solver baseline, report unsuccessful and abstained cases, and prohibit automatic changes to safety-critical operating limits before project-level validation. Such evidence is currently insufficient to rank these proposed paradigms by geothermal field maturity.

Overall, geothermal LLM systems should be designed as auditable decision interfaces to validated physical models rather than autonomous substitutes for governing-equation solvers. Combining task-grounded retrieval, calibrated uncertainty, physics-informed prediction, numerical-simulator feedback and domain-expert approval offers a promising—though largely unvalidated—path toward reliable geothermal engineering support.

5. Conclusions

This narrative and critical review examined the roles, evidence limitations and proposed integration pathways of large language models (LLMs) in geothermal resource development. The principal conclusions are summarized as follows.

(1) The most immediately plausible role of LLMs is the retrieval, extraction and integration of geological and engineering information. Peer-reviewed results from related geoscience, oil-and-gas and energy tasks support the feasibility of several enabling methods, but do not establish equivalent performance on geothermal records. Direct geothermal evaluation, including held-out sites and expert review, remains essential.

(2) LLMs should complement, not replace, physical models in reservoir characterization, drilling-risk assessment and design. Conventional machine-learning and numerical-simulation studies demonstrate capabilities for quantitative prediction, whereas end-to-end LLM-assisted geothermal inversion, early warning and optimization still require task-specific validation. Physical consistency, data provenance and independent verification are critical.

(3) A proposed path toward more capable geothermal information systems integrates domain-restricted retrieval, knowledge-graph reasoning, multimodal representations and validated simulators. The four-layer multimodal roadmap and the LLM–physical-model coupling paradigms synthesized in this review are conceptual frameworks, not deployed or experimentally validated geothermal foundation models. Their utility should be tested against explicitly defined engineering tasks and failure criteria.

(4) Future research should prioritize task-specific geothermal benchmarks with expert-approved annotations, independent-site evaluation, calibrated uncertainty and risk-sensitive thresholds. Transparent reporting of data limitations, model failures, cross-domain transfer gaps, and human-review requirements will be necessary before LLM-based recommendations are adopted in safety-critical geothermal workflows.

Author Contributions: X.L.: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing—review and editing. Z.S.: Data curation, Formal analysis, Visualization, Writing—original draft. T.Z.: Writing—review and editing. S.W.: Writing—review and editing. Z.Y.: Data curation, Software. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Science and Technology Development Plan Project of Jilin Province, China (Project No.: YDZJ202401370ZYTS).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to thank the Science and Technology Development Plan Project of Jilin Province, China (YDZJ202401370ZYTS).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
CPU	Central Processing Unit
EGS	Enhanced Geothermal System
GDR	Geothermal Data Repository
GSHP	Ground Source Heat Pump
HDR	Hot Dry Rock
LLM	Large Language Model
LoRA	Low-Rank Adaptation
MBM	Margin-Based Loss
MPD	Managed Pressure Drilling
NDCG	Normalized Discounted Cumulative Gain
NLP	Natural Language Processing
NREL	National Renewable Energy Laboratory
ORC	Organic Rankine Cycle
R^2	Coefficient of Determination
RAG	Retrieval-Augmented Generation
THMC	Thermal–Hydraulic–Mechanical–Chemical
TFM	Thermal–Fluid–Mechanical

References

1. Horne, R.; Genter, A.; McClure, M.; Ellsworth, W.; Norbeck, J.; Schill, E. Enhanced geothermal systems for clean firm energy generation. *Nat. Rev. Clean Technol.* **2025**, *1*, 148–160. [CrossRef]
2. Li, G.; Xie, Z.; Huang, Z.; Wu, X.; Zou, W.; Sun, Z.; Yang, R.; Cui, Q. Status and challenges of deep geothermal exploitation and energy storage technology: A comprehensive review. *GeoEnergy Commun.* **2025**, *1*, 5. [CrossRef]
3. Ullah, J.; Li, H.; Soupios, P.; Ehsan, M. Optimizing geothermal reservoir modeling: A unified bayesian PSO and BiGRU approach for precise history matching under uncertainty. *Geothermics* **2024**, *119*, 102958. [CrossRef]
4. Han, B.; Susnjak, T.; Mathrani, A. Automating Systematic Literature Reviews with Retrieval-Augmented Generation: A Comprehensive Overview. *Appl. Sci.* **2024**, *14*, 9103. [CrossRef]
5. Teng, T.; Yang, S.; Zhao, Y.; Cui, L.; Zhang, J. Quantitative determination of CH_4 production and CO_2 storage in coal mine goaf: A CO_2-ECBM recovery model that considers zonal permeability. *Green Smart Min. Eng.* **2025**, *2*, 98–108. [CrossRef]
6. Li, X.; Zhu, L.; Zhang, W.; Xiong, F.; Dang, F.; Liu, F.; Cao, Z. Machine Learning-Driven Multi-Scale Modeling and Digital Twin Evolution for Geothermal Reservoirs and Underground Thermal Storage. *Appl. Sci.* **2026**, *16*, 8301. [CrossRef]
7. Nkinyam, C.M.; Ujah, C.O.; Asadu, C.O.; Kallon, D.V.V. Exploring geothermal energy as a sustainable source of energy: A systemic review. *Unconv. Resour.* **2025**, *6*, 100149. [CrossRef]
8. Shah, M.; Prajapati, M.; Yadav, K.; Sircar, A. A comprehensive review of geothermal energy storage: Methods and applications. *J. Energy Storage* **2024**, *98*, 113019. [CrossRef]
9. Lee, K.C. Classification of geothermal resources by exergy. *Geothermics* **2001**, *30*, 431–442. [CrossRef]
10. Wang, H.; Chen, S. Insights into the Application of Machine Learning in Reservoir Engineering: Current Developments and Future Trends. *Energies* **2023**, *16*, 1392. [CrossRef]
11. Magaji, A.; Dou, B.; Gola, G.; Alkhulaidi, G.; Ibrahim, A.-W.; Ngata, M.R. Current status and advancement from high yield and oilfield geothermal energy production: A systematic review. *Appl. Therm. Eng.* **2025**, *258*, 124575. [CrossRef]
12. Elshehabi, T.; Alfehaid, M. Sustainable Geothermal Energy: A Review of Challenges and Opportunities in Deep Wells and Shallow Heat Pumps for Transitioning Professionals. *Energies* **2025**, *18*, 811. [CrossRef]
13. Ahmed, A.A.; Assadi, M.; Kalantar, A.; Sliwa, T.; Sapińska-Śliwa, A. A Critical Review on the Use of Shallow Geothermal Energy Systems for Heating and Cooling Purposes. *Energies* **2022**, *15*, 4281. [CrossRef]
14. Salhein, K.; Salheen, S.A.; Annekaa, A.M.; Hawsawi, M.; Alhawsawi, E.Y.; Kobus, C.J.; Zohdy, M. A Comprehensive Review of Geothermal Heat Pump Systems. *Processes* **2025**, *13*, 2142. [CrossRef]
15. Ahmad, S.; Rizvi, Z.H.; Wuttke, F. Unveiling soil thermal behavior under ultra-high voltage power cable operations. *Sci. Rep.* **2025**, *15*, 7315. [CrossRef]

16. Ahmad, S.; Rizvi, Z.H.; Wuttke, F. Modeling thermal behavior of buried subsea power cables: A review of deterministic, data-driven, and hybrid approaches. *Front. Built Environ.* **2026**, *12*, 1836114. [CrossRef]
17. Vižintin, G.; Šporin, J.; Vukelić, Ž. Geothermal Potential of Pumping–Injecting Doublet Wells. *Energies* **2024**, *17*, 980. [CrossRef]
18. Olasolo, P.; Juárez, M.C.; Morales, M.P.; D'Amico, S.; Liarte, I.A. Enhanced geothermal systems (EGS): A review. *Renew. Sustain. Energy Rev.* **2016**, *56*, 133–144. [CrossRef]
19. Fedus, W.; Zoph, B.; Shazeer, N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *J. Mach. Learn. Res.* **2022**, *23*, 1–39.
20. Kumar, P. Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artif. Intell. Rev.* **2024**, *57*, 260. [CrossRef]
21. Moraga, J.; Duzgun, H.S.; Cavur, M.; Soydan, H. The Geothermal Artificial Intelligence for geothermal exploration. *Renew. Energy* **2022**, *192*, 134–149. [CrossRef]
22. Brown, A.; Roman, M.; Devereux, B. A Systematic Literature Review of Retrieval-Augmented Generation: Techniques, Metrics, and Challenges. *Big Data Cogn. Comput.* **2025**, *9*, 320. [CrossRef]
23. Linders, J.; Tomczak, J.M. Knowledge graph-extended retrieval augmented generation for question answering. *Appl. Intell.* **2025**, *55*, 1102. [CrossRef]
24. Gao, H.; Wu, X.; Liang, L.; Sheng, H.; Si, X.; Gao, H.; Li, Y. A foundation model enpowered by a multi-modal prompt engine for universal seismic geobody interpretation across surveys. *Inf. Fusion* **2026**, *125*, 103437. [CrossRef]
25. Chebbi, A.; Kolade, B. Towards EnergyGPT: A Large Language Model Specialized for the Energy Sector. *IEEE Access* **2026**, *14*, 76920–76964. [CrossRef]
26. Zhou, B.; Li, K. Fusing Geoscience Large Language Models and Lightweight RAG for Enhanced Geological Question Answering. *Geosciences* **2025**, *15*, 382. [CrossRef]
27. Arslan, M.; Mahdjoubi, L.; Munawar, S. Driving sustainable energy transitions with a multi-source RAG-LLM system. *Energy Build.* **2024**, *324*, 114827. [CrossRef]
28. Yan, W.; Wu, J.; Chen, G.; Su, S.; Shen, P.; Wong, D.F.; Zhou, W.-H. Standardization and interpretable analysis of geological database using retrieval-augmented large language model. *Geod. AI* **2026**, *6*, 100047. [CrossRef]
29. Ma, J.; Zhou, Y.; Su, L.; Yang, H.; He, L. A domain-specific language model for engineering-scale geological reasoning and mineral exploration in the Qin-Hang belt. *Eng. Appl. Artif. Intell.* **2026**, *167*, 113651. [CrossRef]
30. Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A.S.; Ceder, G.; Persson, K.A.; Jain, A. Structured information extraction from scientific text with large language models. *Nat. Commun.* **2024**, *15*, 1418. [CrossRef]
31. Lee, J.; Jung, W.; Baek, S. In-House Knowledge Management Using a Large Language Model: Focusing on Technical Specification Documents Review. *Appl. Sci.* **2024**, *14*, 2096. [CrossRef]
32. Ma, Z.; Santos, J.E.; Lackey, G.; Viswanathan, H.; O'Malley, D. Information extraction from historical well records using a large language model. *Sci. Rep.* **2024**, *14*, 31702. [CrossRef]
33. Chen, X.; Lu, T.; Wang, Z. LLM-Align: Utilizing Large Language Models for Entity Alignment in Knowledge Graphs. *Data Intell.* **2026**, *8*, 2026r20. [CrossRef]
34. Hu, J.; Asmuni, H.; Wang, K.; Li, Y. Enhanced Knowledge Graph Completion Based on Structure-Aware and Semantic Fusion Driven by Large Language Models. *Electronics* **2025**, *14*, 4521. [CrossRef]
35. Hu, Q.; Tu, X.; Li, A.; Yao, B. TempQA: An LLM-based framework for temporal knowledge graph question answering. *Knowl.-Based Syst.* **2026**, *333*, 114988. [CrossRef]
36. Neumann, F.; Norden, B.; Balkan-Pazvantoğlu, E.; Elbarbary, S.; Petrunin, A.G.; Elger, K.; Jennings, S.; Dergunova, V.; Fuchs, S. *The* 2024 release of the Global Heat Flow Database (GHFDB): Quality assessment, metadata standards, and a century of geothermal data. *Earth Syst. Sci. Data* **2026**, *18*, 4639–4668. [CrossRef]
37. Suzuki, A.; Shi, S.; Sakai, T.; Fukui, K.; Onodera, S.; Ishizaki, J.; Hashida, T. Automated parameter estimation for geothermal reservoir modeling using machine learning. *Renew. Energy* **2024**, *224*, 120243. [CrossRef]
38. Al-Fakih, A.; Al-khudafi, A.; Koeshidayatullah, A.; Kaka, S.L.; Al-Gathe, A. Forecasting geothermal temperature in western Yemen with Bayesian-optimized machine learning regression models. *Geotherm. Energy* **2025**, *13*, 4. [CrossRef]
39. Qiu, Q.; Tian, M.; Tao, L.; Xie, Z.; Ma, K. Semantic information extraction and search of mineral exploration data using text mining and deep learning methods. *Ore Geol. Rev.* **2024**, *165*, 105863. [CrossRef]
40. Yu, Z.; Ye, X.; Zhang, Y.; Gao, P.; Huang, Y. Experimental research on the thermal conductivity of unsaturated rocks in geothermal engineering. *Energy* **2023**, *282*, 129019. [CrossRef]
41. Okoroafor, E.R.; Smith, C.M.; Ochie, K.I.; Nwosu, C.J.; Gudmundsdottir, H.; Aljubran, M.J. Machine learning in subsurface geothermal energy: Two decades in review. *Geothermics* **2022**, *102*, 102401. [CrossRef]
42. Hatvani, I.G.; Mudelsee, M.; Kern, Z. Essential questions in earth and geosciences according to large language models. *Open Geosci.* **2024**, *16*, 20220677. [CrossRef]

43. Gurina, E.; Klyuchnikov, N.; Antipova, K.; Koroteev, D. Forecasting the abnormal events at well drilling with machine learning. *Appl. Intell.* **2022**, *52*, 9980–9995. [CrossRef]
44. Yi, Z.; Li, Z.; Yi, M.; Wang, D.; Cheng, P. Intelligent Real-Time Risk Evaluation and Drilling Parameter Optimization for Enhanced Safety in Deep-Well Operations. *Processes* **2025**, *13*, 3102. [CrossRef]
45. Park, D.; Moon, H.; Ryu, S. A self-correcting multi-agent LLM framework for language-based physics simulation and explanation. *npj Artif. Intell.* **2026**, *2*, 10. [CrossRef]
46. Arévalo, P.; Jurado, F. Impact of Artificial Intelligence on the Planning and Operation of Distributed Energy Systems in Smart Grids. *Energies* **2024**, *17*, 4501. [CrossRef]
47. Sabbagh, V.B.; Lima, C.B.C.; Xexéo, G. Comparative Analysis of Single and Multiagent Large Language Model Architectures for Domain-Specific Tasks in Well Construction. *SPE J.* **2024**, *29*, 6869–6882. [CrossRef]
48. Liang, C.; Cui, Y.; Shi, R.; Zha, G.; Yin, X.; Xiao, M.; Xu, D.; Duan, X.; Huang, B. GeoAgentic-RAG: A Multi-Agent framework for autonomous geospatial reasoning and visual insight generation with LLM. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *147*, 105195. [CrossRef]
49. Kamalipour, A.A.; Asadi, S.; Amiri Chimeh, M.M. From vectors to knowledge graphs: A comprehensive analysis of modern retrieval-augmented generation architectures. *Comput. Sci. Rev.* **2026**, *61*, 100925. [CrossRef]
50. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; et al. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 39. [CrossRef]
51. Bommasani, R.; Liang, P.; Lee, T. Holistic Evaluation of Language Models. *Ann. N. Y. Acad. Sci.* **2023**, *1525*, 140–146. [CrossRef]
52. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2023**, *55*, 1–38. [CrossRef]
53. Gawlikowski, J.; Tassi, C.R.N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* **2023**, *56*, 1513–1589. [CrossRef]
54. Hüllermeier, E.; Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Mach. Learn.* **2021**, *110*, 457–506. [CrossRef]
55. Zhan, Y.; Yang, R.; You, J.; Huang, M.; Liu, W.; Liu, X. A systematic literature review on incomplete multimodal learning: Techniques and challenges. *Syst. Sci. Control Eng.* **2025**, *13*, 2467083. [CrossRef]
56. Khoee, A.G.; Yu, Y.; Feldt, R. Domain generalization through meta-learning: A survey. *Artif. Intell. Rev.* **2024**, *57*, 285. [CrossRef]
57. Isania, F.; Elahifar, B. Advances in geothermal drilling: A comparative study with oil and gas techniques. *Geoenergy Sci. Eng.* **2025**, *246*, 213638. [CrossRef]
58. Kibikas, W.; Chang, C.; Bauer, S.J.; Nakagawa, S.; Dobson, P.; Kneafsey, T.; Samuel, A. Time-dependent thermal degradation of lost circulation materials in geothermal systems. *Geothermics* **2024**, *121*, 103038. [CrossRef]
59. Song, X.; Li, G.; Huang, Z.; Shi, Y.; Wang, G.; Song, G.; Xu, F. Review of high-temperature geothermal drilling and exploitation technologies. *Gondwana Res.* **2023**, *122*, 315–330. [CrossRef]
60. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [CrossRef]
61. Silva Filho, T.; Song, H.; Perello-Nieto, M.; Santos-Rodriguez, R.; Kull, M.; Flach, P. Classifier calibration: A survey on how to assess and improve predicted class probabilities. *Mach. Learn.* **2023**, *112*, 3211–3260. [CrossRef]
62. Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G.E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229. [CrossRef]
63. Shima, A.; Ishitsuka, K.; Lin, W.; Bjarkason, E.K.; Suzuki, A. Modeling unobserved geothermal structures using a physics-informed neural network with transfer learning of prior knowledge. *Geotherm. Energy* **2024**, *12*, 38. [CrossRef]
64. Xue, Z.; Bi, J.; Ma, H.; Sun, Z.; Chen, Z. A robust physics-constrained neural operator framework for efficient geothermal resource development. *Nat. Commun.* **2026**, *17*, 6929. [CrossRef]
65. 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. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
66. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
67. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.* **2022**, *55*, 1–37. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.