Physics-Informed Machine Learning for Geological CO_2 Storage Optimization

Mahendra Soni

Integral University, Lucknow

Abstract

Geological CO_2 storage (GCS) is increasingly recognized as an indispensable pillar of net-zero emissions strategies, yet optimizing injection design, storage capacity, and containment safety across heterogeneous subsurface formations remains computationally and technically demanding. Conventional numerical reservoir simulation, while physically rigorous, is too computationally expensive to support the exhaustive scenario evaluation and real-time decision-making that large-scale, multi-site GCS deployment requires. Purely data-driven machine learning models, by contrast, can be fast but often violate fundamental physical laws when extrapolated beyond their training distribution, undermining confidence in high-consequence subsurface applications. Physics-informed machine learning (PIML) has emerged as a synthesis of these two paradigms, embedding governing equations—mass and energy conservation, Darcy flow, geomechanical equilibrium—directly into neural network architectures and loss functions to produce models that are both computationally efficient and physically consistent. This review examines the theoretical foundations, architectures, and field applications of PIML for GCS optimization, drawing on recent case studies from heterogeneous depleted gas reservoirs, geothermal-analogue reservoir characterization, and geomechanically informed wellbore stability assessment. It further considers methodological parallels with large-scale AI-driven optimization frameworks developed in public health and healthcare administration, which offer transferable lessons on deploying physics- and rule-constrained AI systems at national scale. The review concludes with a discussion of persistent challenges in training stability, multiphysics coupling, and uncertainty quantification, and proposes a research agenda for advancing PIML toward operational deployment in commercial-scale CO_2 storage projects.

Keywords: physics-informed machine learning, geological CO_2 storage, neural operators, reservoir optimization, geomechanics, reactive transport, deep learning

Introduction

The scale-up of carbon capture, utilization, and storage (CCUS) required to meet global climate targets demands the injection and permanent containment of gigatonnes of CO_2 annually within depleted hydrocarbon reservoirs, deep saline aquifers, and other subsurface formations. Achieving this scale safely and cost-effectively requires optimization across a vast decision space—well placement, injection rate scheduling, and reservoir management strategy—all while respecting hard physical constraints on pressure buildup, plume containment, and geomechanical stability. Traditional reservoir simulation tools, grounded in coupled partial differential equations governing multiphase flow, geomechanics, and reactive transport, remain the gold standard for predictive accuracy but are prohibitively expensive to run across the thousands of geologic realizations and operational scenarios that robust optimization under uncertainty demands.

Machine learning has been proposed as a means of accelerating this optimization process, with deep neural networks trained as surrogate models capable of approximating expensive simulator outputs at a fraction of the computational cost. However, purely data-driven surrogates trained without physical constraints are prone to non-physical extrapolation, poor generalization outside their training distribution, and a lack of interpretability that undermines regulatory confidence—concerns that are particularly acute in GCS, where erroneous predictions can translate into containment failure or wasted storage capacity (Williams et al., 2026). Physics-informed machine learning (PIML) addresses these shortcomings by embedding governing physical laws directly into the learning process, either through soft penalty terms in the loss function (as in physics-informed neural networks, or PINNs) or through architectural constraints that guarantee conservation properties by construction (Raissi et al., 2019).

This review synthesizes recent advances in PIML for GCS optimization, integrating insights from heterogeneous reservoir case studies (Williams et al., 2026), geothermal-analogue reservoir characterization (Yeboah et al., 2026), and machine-learning-enhanced geomechanical risk assessment (Akpabli et al., 2026), alongside the growing PIML literature applied directly to CO_2 transport, storage capacity prediction, and coupled flow-geomechanics modeling (Du et al., 2023; Shokouhi et al., 2021; Tang et al., 2022; Wen et al., 2023; Yan et al., 2024). It further draws methodological parallels with large-scale, constraint-aware AI optimization frameworks developed for public health and healthcare administration (Nguyen, 2026a, 2026b, 2026c), which offer transferable lessons on deploying physically and institutionally constrained AI systems at national scale. The objective is to provide a comprehensive, integrative synthesis for researchers and practitioners seeking to advance PIML from research prototype toward operational deployment in commercial CO_2 storage projects.

The Optimization Challenge in Geological CO_2 Storage

Decision Variables and Physical Constraints

Optimizing a GCS project requires simultaneously resolving decisions on injector well count, placement, and completion interval; injection rate and pressure scheduling over the project lifetime; and monitoring network design—all subject to constraints on maximum allowable bottom-hole pressure, fracture gradient, and fault reactivation stress thresholds. Because these constraints emerge from coupled multiphysics processes—multiphase flow, geomechanical stress redistribution, and geochemical alteration—optimization cannot proceed on flow considerations alone; geomechanical risk must be evaluated jointly with storage efficiency (Akpabli et al., 2026).

Heterogeneity and the Curse of Dimensionality

Field-scale reservoir heterogeneity, as demonstrated in the Evetts site case study in the depleted-gas Permian Basin, introduces substantial uncertainty into predictions of CO_2 plume migration, pressure buildup, and ultimate storage capacity (Williams et al., 2026). Robust optimization under this uncertainty requires evaluating outcomes across an ensemble of geologically plausible realizations—often numbering in the hundreds or thousands—a computational burden that renders full-physics simulation-in-the-loop optimization impractical for most projects without surrogate modeling.

The Case for Physics-Informed Approaches

Data-driven surrogate models trained purely on simulation output can dramatically reduce computational cost, but their predictions are only as reliable as the training data's coverage of the operational and geologic scenario space. Because CO_2 storage optimization inherently involves searching beyond previously simulated scenarios—new well configurations, injection strategies, or geologic settings—surrogate models risk extrapolating into physically inconsistent regimes. PIML directly addresses this limitation by constraining model outputs to satisfy governing conservation laws, improving generalization and providing a principled basis for trust in model predictions used to guide multimillion-dollar injection decisions.

Foundations of Physics-Informed Machine Learning

Physics-Informed Neural Networks (PINNs)

The foundational PINN framework, introduced by Raissi et al. (2019), incorporates the residuals of governing partial differential equations directly into the neural network's training loss function alongside data-fitting terms, enabling the network to learn solutions that are simultaneously consistent with observed data and the underlying physics. This approach has since been extended to a wide range of subsurface flow and transport problems relevant to GCS. Du et al. (2023) applied a Fourier-embedded PINN architecture to model density-driven convective mixing of CO_2 in brine—a slow but geologically important process that enhances long-term storage security through solubility trapping—demonstrating that PINNs enriched with frequency-domain feature embeddings can capture the fine-scale fingering instabilities characteristic of this process while remaining consistent with the underlying Boussinesq flow equations.

Physics-Informed Convolutional and U-Net Architectures

Beyond meshless PINN formulations, physics-informed convolutional architectures embed physical constraints into spatially structured deep learning models. Shokouhi et al. (2021) developed a physics-informed U-Net architecture that couples the spatial pattern-recognition capacity of convolutional neural networks with explicit incorporation of governing flow equations to predict CO_2 storage site response—pressure and saturation evolution—demonstrating improved consistency and reliability relative to purely data-driven convolutional surrogates, particularly under scenarios involving limited or noisy training data.

Neural Operators for Rapid, Resolution-Independent Prediction

A parallel line of research has developed neural operator architectures capable of learning mappings between function spaces—for example, from permeability fields to pressure and saturation fields—independent of the specific spatial discretization used during training. Wen et al. (2023) demonstrated a nested Fourier neural operator capable of producing real-time, high-resolution predictions of CO_2 plume migration and pressure evolution across large-scale, heterogeneous geomodels, achieving computational speedups of several orders of magnitude relative to conventional reservoir simulation while preserving strong agreement with high-fidelity results. Such neural operator approaches are particularly well suited to field-scale optimization workflows, such as those required for heterogeneous reservoirs like the Evetts site (Williams et al., 2026), where thousands of injection strategy evaluations across variable reservoir realizations are needed to identify optimal, risk-managed CO_2 storage configurations.

Coupled Flow-Geomechanics Surrogates

Because CO_2 injection-induced pressure buildup can alter the in-situ stress state and reactivate faults, coupling geomechanical prediction with flow prediction is essential to any physically grounded optimization workflow. Tang et al. (2022) developed a deep-learning-based coupled flow-geomechanics surrogate model trained to jointly predict pressure, saturation, and geomechanical deformation during CO_2 sequestration, directly capturing the stress-dependent permeability feedback that governs both injectivity evolution and fault reactivation risk—capabilities that parallel the machine-learning-enhanced geomechanical characterization and wellbore stability workflows demonstrated in offshore reservoir settings (Akpabli et al., 2026).

Applications to Reservoir and Site Characterization

Petrophysical Property Prediction as a Physics-Informed Input

Accurate characterization of porosity, permeability, and mineralogical composition is foundational to any physics-informed optimization workflow, since PIML models are only as reliable as the geologic property fields on which their physical constraints are evaluated. Machine learning-enhanced petrophysical workflows applied to the Three Forks geothermal reservoir in the Williston Basin illustrate how well log, core, and seismic attribute data can be integrated to predict reservoir quality at higher spatial resolution than conventional deterministic interpolation methods (Yeboah et al., 2026). These high-resolution property fields are directly usable as inputs to physics-informed surrogate models of CO_2 storage capacity and injectivity, improving the fidelity of downstream optimization.

Geomechanical Constraint Integration

Physics-informed optimization of CO_2 storage must explicitly incorporate geomechanical risk constraints—maximum sustainable injection pressure, fault slip potential, and wellbore stability—rather than treating them as post-hoc checks on an optimized injection schedule. The machine-learning-enhanced geomechanical characterization and wellbore stability assessment demonstrated in the offshore South Tano Field illustrates how stress field estimation, rock strength prediction, and drilling parameter analysis can be integrated into a proactive risk-management workflow (Akpabli et al., 2026); embedding this class of geomechanical model directly within a physics-informed optimization loop—rather than as an external constraint check—

represents a natural and valuable extension of current PIML architectures for CO_2 storage.

Transfer of Physics-Informed Methods from Geothermal Systems

Geothermal reservoirs provide a valuable proving ground for physics-informed methods relevant to CO_2 storage, given their shared reliance on coupled thermal-hydraulic-mechanical process modeling. Yan et al. (2024) demonstrated a physics-informed machine learning framework for noniterative optimization in geothermal energy recovery, embedding heat and mass transport physics directly into the optimization loop to avoid the repeated forward-simulation calls required by conventional iterative optimization schemes. This noniterative optimization paradigm—learning an optimized decision variable field directly rather than iteratively refining candidate solutions through repeated physics-informed surrogate evaluations—represents a promising direction for accelerating CO_2 storage optimization workflows involving well placement and injection scheduling under heterogeneous reservoir conditions such as those characterized at the Evetts site (Williams et al., 2026).

Physics-Informed Optimization Workflows for CO_2 Storage

Surrogate-Assisted Optimization Under Uncertainty

The central value proposition of PIML for GCS optimization lies in enabling surrogate-assisted optimization under geologic and operational uncertainty. Rather than optimizing against a single deterministic reservoir model, physics-informed surrogates allow rapid evaluation of injection strategies across large ensembles of geologically plausible realizations, supporting robust optimization formulations that explicitly account for the heterogeneity documented in field case studies such as the Evetts site (Williams et al., 2026). Because the surrogate is physically constrained, the resulting optimized strategies retain higher fidelity to true reservoir behavior than would be achievable with purely data-driven surrogates trained on a necessarily incomplete simulation ensemble.

Multi-Objective Formulations

CO_2 storage optimization is inherently multi-objective, requiring simultaneous consideration of storage capacity maximization, injection cost minimization, and geomechanical and containment risk minimization. Physics-informed surrogate models—by directly encoding the governing flow and geomechanical equations—provide a principled basis for evaluating these competing objectives within a unified optimization framework, avoiding the need for separate, potentially inconsistent surrogate models for each objective.

Real-Time Optimization and Adaptive Management

Beyond pre-injection design optimization, physics-informed neural operator architectures such as those developed by Wen et al. (2023) support real-time, adaptive injection management, in which observed monitoring data—bottom-hole pressure, seismic attributes, or geochemical sampling—can be rapidly assimilated to update reservoir predictions and adjust injection strategy dynamically over the project lifetime. This capability moves CO_2 storage management from a static, pre-optimized plan toward a continuously adaptive, physics-consistent decision-support system.

Cross-Domain Lessons: Physics- and Rule-Constrained AI at National Scale

While the technical foundations of PIML for GCS optimization are rooted in subsurface physics, valuable operational and governance lessons can be drawn from large-scale AI optimization frameworks developed in other high-stakes, constraint-bound domains. National-scale AI frameworks developed for chronic disease prevention and early intervention illustrate how predictive models can be constrained by clinical guidelines and public health protocols—an operational analogue to embedding physical conservation laws into subsurface prediction models—ensuring that AI-driven recommendations remain consistent with domain-established rules even when applied to novel patient populations outside the original training distribution (Nguyen, 2026a). This principle of constraint-consistent generalization is directly relevant to PIML in GCS, where optimized injection strategies must remain physically valid even for geologic and operational scenarios not directly represented in the training simulation ensemble.

Similarly, frameworks for transforming prior authorization through artificial intelligence demonstrate how AI systems can be designed to operate within a fixed set of institutional and regulatory constraints—reducing administrative burden while ensuring that automated decisions remain auditable and compliant with policy rules (Nguyen, 2026b). This parallels the regulatory environment surrounding GCS, where physics-informed optimization outputs must be auditable against established containment and pressure-management thresholds to satisfy monitoring, verification, and accounting (MVA) requirements; a PIML system whose outputs are explicitly grounded in governing physical equations offers a more defensible basis for regulatory approval than an unconstrained data-driven alternative.

Economic impact analyses of AI adoption in healthcare—quantifying national cost savings and productivity gains attributable to constrained, guideline-consistent AI deployment—offer a further template for justifying investment in physics-informed optimization infrastructure for CCUS (Nguyen, 2026c). Just as healthcare AI systems that respect clinical constraints have demonstrated measurable cost savings through more reliable, guideline-consistent decision support, PIML-based CO_2 storage optimization systems can be expected to reduce costs associated with over-conservative injection design, unnecessary monitoring infrastructure, or, conversely, containment failures arising from unconstrained model extrapolation—an economic case that merits explicit quantification as PIML systems mature toward operational deployment.

Collectively, these cross-domain parallels reinforce a broader principle: the value of physics- or rule-informed constraints in machine learning is not unique to subsurface engineering, but reflects a general requirement for trustworthy AI deployment in domains where erroneous extrapolation carries substantial financial, safety, or public-health consequences. The maturation of PIML for GCS optimization can accordingly benefit from governance and validation frameworks developed in these adjacent, already-scaled AI application domains.

Challenges and Limitations

Despite substantial progress, several challenges continue to constrain the operational deployment of PIML for GCS optimization.

Training instability and loss function balancing. PINN training often requires careful balancing of data-fitting and physics-residual loss terms, and can suffer from stiff gradient pathologies that impede convergence, particularly for the highly nonlinear, multiphase flow equations relevant to CO_2 injection. Improved training algorithms and adaptive loss-weighting schemes remain an active area of methodological research.

Multiphysics coupling complexity. Fully coupled flow-geomechanics-geochemistry PIML models capable of jointly resolving pressure evolution, stress redistribution, and mineral reaction kinetics remain computationally and architecturally demanding to construct, with most current PIML applications addressing only a subset of the relevant coupled physics.

Uncertainty quantification. While PIML models improve physical consistency relative to purely data-driven surrogates, quantifying predictive uncertainty—particularly epistemic uncertainty arising from limited training data coverage—remains an open challenge essential to supporting risk-informed decision-making in GCS optimization.

Generalization across geologic settings. Physics-informed constraints improve generalization relative to unconstrained models, but PIML surrogates trained on one geologic setting, such as a depleted gas reservoir, may still require substantial retraining or transfer learning to perform reliably in structurally distinct settings such as saline aquifers or fractured basement reservoirs.

Regulatory and industry validation. Translating PIML research prototypes into industry-accepted decision-support tools requires extensive validation against field data and established simulation benchmarks, as well as the development of standardized verification protocols that regulators and operators can trust.

Future Directions

Future research should prioritize the development of fully coupled, multiphysics PIML architectures capable of jointly resolving flow, geomechanics, and reactive transport within a single physically consistent optimization framework, building on the individual-process advances reviewed here (Du et al., 2023; Shokouhi et al., 2021; Tang et al., 2022; Wen et al., 2023). Integration of high-resolution petrophysical and geomechanical characterization workflows—such as those demonstrated for the Three Forks geothermal reservoir (Yeboah et al., 2026) and the South Tano offshore field (Akpabli et al., 2026)—directly into PIML training pipelines represents a promising avenue for improving the fidelity of site-specific optimization.

Continued methodological transfer from noniterative, physics-informed optimization approaches developed in geothermal energy systems (Yan et al., 2024) could substantially accelerate CO_2 storage optimization workflows, particularly for heterogeneous, multi-well injection scenarios such as those characterized at the Evetts site (Williams et al., 2026). Finally, adopting governance and validation frameworks informed by constraint-consistent AI deployment in public health and healthcare administration (Nguyen, 2026a, 2026b, 2026c) may help establish the auditability and regulatory trust required for PIML systems to transition from research demonstration to routine use in commercial-scale CO_2 storage project design and operation.

Conclusion

Physics-informed machine learning represents a critical methodological advance for geological CO_2 storage optimization, resolving the tension between the computational efficiency of data-driven surrogate models and the physical fidelity required for high-consequence subsurface decision-making. Applications spanning density-driven convective trapping, coupled flow-geomechanics prediction, real-time neural operator-based monitoring, and petrophysical and geomechanical site characterization collectively demonstrate that embedding governing physical laws into machine learning architectures substantially improves both computational tractability and physical reliability relative to purely data-driven or purely physics-based approaches alone. Cross-domain lessons from constraint-consistent AI deployment in public health and healthcare administration further suggest a pathway toward the governance and validation frameworks necessary for PIML systems to achieve regulatory trust and operational adoption. Realizing the full potential of physics-informed machine learning in this domain will require sustained investment in multiphysics coupling, training stability, and uncertainty quantification—investments essential to ensuring that geological CO_2 storage can be optimized safely, economically, and at the scale required to meet global climate objectives.

References

1. Du, H., Zhao, Z., Cheng, H., Yan, J., & He, Q. (2023). Modeling density-driven flow in porous media by physics-informed neural networks for CO2 sequestration. *Computers and Geotechnics, 159*, 105433. https://doi.org/10.1016/j.compgeo.2023.105433
2. Nguyen, T. T. (2026a). AI-powered precision public health: A national framework for targeting chronic disease prevention and early intervention. *International Journal of Emerging Trends in Computer Science and Information Technology, 7*(3), 10–20.
3. Nguyen, T. T. (2026b). Transforming prior authorization through artificial intelligence: A national framework for reducing administrative burden and improving patient access. *International Journal of AI, BigData, Computational and Management Studies, 7*(3), 17–27.
4. Nguyen, T. T. (2026c). The economic impact of AI adoption in healthcare: Estimating national cost savings, productivity gains, and long-term health outcomes. *International Journal of Artificial Intelligence, Data Science, and Machine Learning, 7*(3), 1–11.
5. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational*

Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

6. Shokouhi, P., Kumar, V., Prathipati, S., Hosseini, S. A., Giles, C. L., & Kifer, D. (2021). Physics-informed deep learning for prediction of CO2 storage site response. *Journal of Contaminant Hydrology, 241*, 103835. https://doi.org/10.1016/j.jconhyd.2021.103835
7. Tang, M., Ju, X., & Durlofsky, L. J. (2022). Deep-learning-based coupled flow-geomechanics surrogate model for CO2 sequestration. *International Journal of Greenhouse Gas Control, 118*, 103692. https://doi.org/10.1016/j.ijggc.2022.103692
8. Wen, G., Li, Z., Long, Q., Azizzadenesheli, K., Anandkumar, A., & Benson, S. M. (2023). Real-time high-resolution CO2 geological storage prediction using nested Fourier neural operators. *Energy & Environmental Science, 16*(4), 1732–1741. https://doi.org/10.1039/D2EE04204E
9. Yan, B., Gudala, M., Hoteit, H., Sun, S., Wang, W., & Jiang, L. (2024). Physics-informed machine learning for noniterative optimization in geothermal energy recovery. *Applied Energy, 365*, 123179. https://doi.org/10.1016/j.apenergy.2024.123179