AI-Driven Uncertainty Quantification for Long-Term CO_2 Storage Security

Ajay Kandpal[1], Ramesh Kandari[2]

[1]Associate Professor, Department of Computer Sciences, Uttarakhand Technical University, Dehradun, UK

[2]Research Scholar, Department of Environmental Sciences, Uttarakhand Technical University, Dehradun, UK

ABSTRACT

Long-term security of geological CO_2 storage (GCS) depends fundamentally on the ability to characterize and communicate uncertainty in predictions of plume migration, pressure evolution, and containment integrity across decadal to centennial project timescales. Point-estimate predictions from deterministic reservoir simulations or single-output machine learning models, however accurate on average, provide an insufficient basis for the risk-informed regulatory and operational decisions that long-term storage security demands, particularly given the deep geologic uncertainty inherent to subsurface characterization and the practical impossibility of directly validating century-scale containment predictions within a human planning horizon. Artificial intelligence-driven uncertainty quantification (UQ)—spanning Bayesian neural networks, ensemble-based deep learning surrogates, and simulation-based inference techniques—has emerged as an essential complement to point-prediction machine learning, enabling the propagation of both aleatoric uncertainty arising from inherent geologic variability and epistemic uncertainty arising from limited data and model knowledge into calibrated, decision-relevant probabilistic forecasts. This review examines the theoretical foundations and applications of AI-driven UQ for long-term CO_2 storage security, synthesizing recent advances in dimension-adaptive Bayesian neural networks, deep convolutional encoder-decoder surrogates, AI-enhanced data assimilation, and uncertainty-aware digital shadow frameworks. It draws on case studies from heterogeneous depleted gas reservoirs, geothermal-analogue petrophysical characterization, and geomechanically informed wellbore stability assessment, and further considers methodological parallels with large-scale AI-driven predictive analytics and economic impact frameworks developed for national healthcare systems. The review concludes by identifying persistent challenges in uncertainty calibration, computational scalability to three-dimensional and century-scale problems, and the translation of quantified uncertainty into regulatory decision criteria, and proposes a research agenda for advancing AI-driven UQ toward routine use in long-term CO_2 storage security assessment.

Keywords: uncertainty quantification, artificial intelligence, geological CO_2 storage, Bayesian neural networks, deep learning, long-term storage security, epistemic uncertainty

1. Introduction

Geological CO_2 storage (GCS) is fundamentally a long-duration risk management undertaking: injected CO_2 must remain securely contained not merely for the years or decades of active injection but for centuries thereafter, over timescales that vastly exceed the direct observational record available to validate any given predictive model. This temporal mismatch between the planning horizon required for storage security assurance and the monitoring history actually available creates a central epistemic challenge for GCS: predictions of plume migration, pressure evolution, and containment integrity must be trusted decades or centuries into the future, based on models trained and validated against comparatively brief periods of observation and a necessarily incomplete characterization of subsurface geology.

Uncertainty quantification (UQ) addresses this challenge by explicitly characterizing the confidence—or lack thereof—associated with any given predictive output, decomposing total predictive uncertainty into aleatoric components arising from inherent, irreducible geologic variability and epistemic components arising from limited data, incomplete model knowledge, or unresolved parameter values that could, in principle, be further constrained with additional information. For regulatory decision-making, this distinction matters: probabilistic containment metrics—for example, the probability that a CO_2 plume exceeds a defined area of review within a specified time horizon—are increasingly required by regulatory frameworks in place of single deterministic point estimates, and artificial intelligence-driven UQ methods are increasingly recognized as essential infrastructure for generating these probabilistic metrics at the computational scale that long-term, uncertainty-quantified storage security assessment demands.

This review synthesizes recent advances in AI-driven UQ for long-term CO_2 storage security, integrating insights from heterogeneous depleted gas reservoir case studies, geothermal-analogue petrophysical characterization (Yeboah et al., 2026), and machine-learning-enhanced geomechanical risk assessment (Akpabli et al., 2026), alongside the growing body of UQ-focused machine learning literature applied directly to GCS (Gahlot et al., 2025; Mahjour et al., 2025; Mo et al., 2019; Seabra et al., 2024). It further draws methodological parallels with large-scale, uncertainty-aware AI predictive analytics and economic impact frameworks developed for public health and healthcare systems (Nguyen, 2026a, 2026c, 2026d), which offer transferable lessons on communicating and acting upon probabilistic risk assessments at scale. The objective is to provide a comprehensive synthesis for researchers, engineers, and regulators seeking to advance AI-driven UQ from research demonstration toward standard practice in long-term CO_2 storage security assessment.

2. The Long-Term Storage Security Challenge

2.1 Why Point Estimates Are Insufficient

A deterministic prediction of plume extent or pressure buildup at a given future time, however accurate on average across an ensemble of scenarios, provides no information about the confidence that should be placed in that specific prediction for a specific site. Long-term storage security decisions—monitoring network design, financial assurance requirements, liability transfer timing—require not a single best estimate but a calibrated probability distribution over possible future outcomes, allowing regulators and operators to reason explicitly about low-probability, high-consequence containment failure scenarios that a point estimate would necessarily obscure.

2.2 Aleatoric and Epistemic Uncertainty in GCS

Aleatoric uncertainty in GCS arises from the inherent, irreducible spatial variability of subsurface geologic properties—heterogeneous porosity and permeability fields that cannot be fully characterized even with extensive site investigation—while epistemic uncertainty arises from limited training data, unresolved model parameters, or incomplete physical process representation that could, in principle, be reduced through additional data collection or model refinement. Explicitly decomposing total predictive uncertainty into these components is valuable because it clarifies where additional monitoring or characterization investment would meaningfully reduce uncertainty (epistemic) versus where uncertainty reflects an irreducible property of the geologic system itself (aleatoric).

2.3 The Computational Burden of Conventional UQ

Conventional approaches to UQ in subsurface flow and transport—Monte Carlo simulation across large ensembles of geologic realizations—provide statistically rigorous uncertainty estimates but require numerous full-physics forward simulations, a computational burden that becomes prohibitive for three-dimensional, century-scale storage security assessments, particularly when uncertainty must be propagated through coupled flow, geomechanical, and geochemical processes simultaneously (Mahjour et al., 2025).

3. AI-Driven UQ Architectures for CO_2 Storage

3.1 Bayesian Neural Networks and Dimension-Adaptive Architectures

Bayesian neural networks extend conventional deep learning architectures by placing probability distributions over network weights rather than point estimates, enabling the network to output calibrated predictive uncertainty alongside point predictions. Mahjour et al. (2025) developed a dimension-adaptive Bayesian neural network architecture for quantifying uncertainty across both two-dimensional and three-dimensional carbon storage models, explicitly decomposing predictive uncertainty into aleatoric components reflecting inherent geological variability and epistemic components reflecting model limitations, and demonstrating that this dimension-adaptive approach enables efficient knowledge transfer between dimensional representations while substantially reducing the computational cost relative to conventional Monte Carlo-based UQ approaches—a capability directly relevant to the three-dimensional, computationally demanding nature of realistic storage security assessments.

3.2 Deep Convolutional Encoder-Decoder Surrogates for Uncertainty Propagation

Deep convolutional encoder-decoder architectures offer a complementary approach to UQ, learning a surrogate mapping from high-dimensional, spatially heterogeneous permeability fields to output pressure and saturation fields that can then be evaluated rapidly across large ensembles of geologic realizations to characterize predictive uncertainty. Mo et al. (2019) developed a deep convolutional encoder-decoder network methodology explicitly formulated as an image-to-image regression strategy, addressing the curse of dimensionality inherent to high-dimensional permeability input fields as well as the saturation discontinuity arising from capillary effects, and demonstrating that the resulting surrogate could support efficient uncertainty quantification of dynamic multiphase flow in heterogeneous media relevant to geological carbon storage—establishing a foundational architecture that subsequent UQ-focused surrogate modeling efforts in the GCS literature have continued to build upon.

3.3 AI-Enhanced Data Assimilation for Continuously Updated Uncertainty

Beyond static, pre-injection uncertainty quantification, AI-enhanced data assimilation techniques enable continuous updating of uncertainty estimates as new monitoring data becomes available over the project lifetime. Seabra et al. (2024) investigated the integration of machine learning surrogate models—Fourier neural operators and Transformer U-Net architectures—with ensemble-based data assimilation techniques, introducing a surrogate-based hybrid Ensemble Smoother with Multiple Data Assimilation (SH-ESMDA) capable of substantially accelerating the data assimilation process relative to conventional approaches while maintaining high-fidelity posterior uncertainty estimates, and further demonstrating that a surrogate-based hybrid randomized maximum likelihood approach could achieve improved uncertainty quantification relative to conventional ensemble smoother methods. This capability for continuously updated, AI-accelerated data assimilation is directly relevant to long-term storage security assessment, where uncertainty estimates generated at the pre-injection design stage must be progressively refined as decades of monitoring data accumulate.

3.4 Uncertainty-Aware Digital Shadows for Long-Term Monitoring

Simulation-based inference techniques provide a further avenue for AI-driven UQ, enabling the characterization of posterior uncertainty in subsurface state conditioned on multimodal monitoring data without requiring explicit likelihood evaluation. Gahlot et al. (2025) developed an uncertainty-aware digital shadow explicitly designed for long-term monitoring of CO_2 injection projects spanning multiple decades, combining simulation-based inference with a novel neural adaptation of nonlinear ensemble filtering; because direct validation of century-scale containment predictions against real long-term monitoring data remains impossible given the field's youth, the framework's performance was instead validated through controlled, in silico computational studies—an approach the authors note is methodologically analogous to the validation strategies employed in long-term climate modeling, where direct empirical validation over the full prediction horizon is similarly infeasible.

4. Integrating Uncertainty Quantification with Site Characterization

4.1 Petrophysical Characterization as an Uncertainty Source and Constraint

The fidelity of any UQ framework for CO_2 storage security is fundamentally bounded by the quality and spatial resolution of the petrophysical characterization on which its geologic input distributions are based. 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 characterize reservoir quality at high spatial resolution while explicitly quantifying the residual uncertainty in porosity and permeability predictions (Yeboah et al., 2026); this residual petrophysical uncertainty constitutes a direct input to the aleatoric and epistemic uncertainty decomposition central to Bayesian neural network and encoder-decoder surrogate UQ frameworks (Mahjour et al., 2025; Mo et al., 2019).

4.2 Geomechanical Uncertainty and Long-Term Stability Risk

Long-term storage security depends not only on flow and containment uncertainty but on geomechanical stability uncertainty—the probability that accumulated pressure and stress redistribution over decades of injection and post-injection monitoring could trigger fault reactivation or wellbore integrity loss. Machine-learning-enhanced geomechanical characterization and wellbore stability assessment, as demonstrated in the offshore South Tano Field, integrates stress field estimation, rock strength prediction, and drilling parameter analysis into a proactive risk-management workflow (Akpabli et al., 2026); embedding the resulting geomechanical risk estimates within a formal AI-driven UQ framework—propagating geomechanical parameter uncertainty through to long-term stability risk estimates—represents an important and still-developing integration point for comprehensive storage security assessment.

4.3 Heterogeneous Reservoir Uncertainty at the Evetts Site

Heterogeneous depleted gas reservoir storage optimization, as exemplified by the Evetts site case study in the Permian Basin, illustrates the practical necessity of uncertainty-quantified assessment at sites where substantial spatial variability in porosity, permeability, and residual saturation directly governs the confidence with which long-term plume migration and pressure evolution can be predicted. AI-driven UQ frameworks such as the dimension-adaptive Bayesian neural network approach of Mahjour et al. (2025) are directly applicable to characterizing this class of heterogeneity-driven uncertainty at computational cost substantially lower than exhaustive Monte Carlo simulation across the full ensemble of plausible geologic realizations.

5. From Quantified Uncertainty to Decision-Relevant Security Metrics

5.1 Probabilistic Containment Metrics

Regulatory frameworks increasingly require probabilistic containment metrics—for example, the probability that a CO_2 plume exceeds a defined area of review within a specified time horizon—rather than deterministic point estimates of plume extent. AI-driven UQ methods, by producing calibrated predictive distributions rather than single-point forecasts, directly enable the construction of these probabilistic metrics, transforming raw uncertainty quantification output into the kind of decision-relevant risk statement that regulatory review processes require.

5.2 Continuously Updated Security Assessment

Because long-term storage security assessment must be revisited and refined as monitoring data accumulates over the project lifetime, AI-enhanced data assimilation frameworks capable of continuously updating uncertainty estimates (Seabra et al., 2024) offer a pathway toward dynamic, rather than static, security assessment—one in which the confidence associated with containment predictions is progressively sharpened as epistemic uncertainty is reduced through accumulated field observation, while aleatoric uncertainty attributable to genuine geologic heterogeneity remains appropriately represented in the security assessment even as monitoring data accumulates.

5.3 Communicating Uncertainty to Regulators and Stakeholders

The translation of quantified uncertainty into a form that regulators, financial institutions, and affected communities can interpret and act upon remains a nontrivial challenge distinct from the underlying technical UQ methodology itself. Uncertainty-aware digital shadow frameworks explicitly designed for long-term monitoring communication (Gahlot et al., 2025) represent an important step toward operationalizing AI-driven UQ output as an ongoing, updatable basis for stakeholder communication regarding storage security, rather than a one-time technical exercise conducted only at the pre-injection design stage.

6. Cross-Domain Lessons: Uncertainty-Aware AI for Long-Term Risk Assessment at National Scale

While the technical foundations of AI-driven UQ for GCS are rooted in subsurface geoscience and Bayesian machine learning, valuable lessons can be drawn from large-scale, uncertainty-aware predictive AI frameworks developed for long-horizon risk assessment in other high-stakes domains. National-scale AI frameworks developed for precision public health illustrate how predictive models can be designed to communicate calibrated confidence in risk assessments over extended time horizons, supporting early intervention decisions that must be made under genuine uncertainty about long-term disease trajectories (Nguyen, 2026a). This principle—that long-horizon risk predictions must be accompanied by explicit, calibrated uncertainty rather than presented as unqualified point forecasts—maps directly onto long-term CO_2 storage security assessment, where containment predictions spanning decades to centuries similarly require explicit uncertainty communication rather than unqualified deterministic forecasts.

National-scale predictive analytics frameworks applied to Medicare and Medicaid populations further demonstrate how AI-driven risk stratification under uncertainty can inform proactive, rather than purely reactive, resource allocation decisions across large, heterogeneous populations monitored over extended periods (Nguyen, 2026d). This proactive, uncertainty-informed resource allocation logic parallels the long-term monitoring network design challenge central to CO_2 storage security, where AI-driven UQ estimates of where containment uncertainty is highest can directly inform the prioritization of monitoring resources toward the storage complex zones and time horizons of greatest concern.

Economic impact analyses of AI adoption in healthcare, quantifying national-scale cost savings attributable to more efficient, uncertainty-informed resource allocation relative to uniform, undifferentiated approaches, offer a further template for constructing the economic case for AI-driven UQ investment in long-term CO_2 storage security assessment (Nguyen, 2026c). Just as healthcare systems that explicitly quantify and act upon predictive uncertainty have demonstrated measurable efficiency gains relative to systems relying on unqualified point predictions, AI-driven UQ for CO_2 storage can be expected to reduce the costs associated with either excessive, undifferentiated precautionary monitoring across an entire storage complex or, conversely, underinvestment in monitoring at genuinely high-uncertainty locations—an economic argument that merits explicit quantification as AI-driven UQ methods mature toward standard practice in commercial-scale CO_2 storage security assessment.

Collectively, these cross-domain parallels reinforce a broader principle: the value of explicit, AI-driven uncertainty quantification for informing long-horizon, high-consequence resource allocation and risk management decisions is not unique to subsurface CO_2 storage, but reflects a general pattern observed across domains where predictions must be trusted and acted upon over extended timescales that preclude direct, complete empirical validation. The maturation of AI-driven UQ for long-term GCS security can accordingly draw on governance, communication, and cost-justification frameworks already developed in these adjacent, more mature large-scale AI application domains.

7. Challenges and Limitations

Despite substantial progress, several challenges continue to constrain the operational deployment of AI-driven UQ for long-term CO_2 storage security.

Uncertainty calibration. A predictive model that outputs uncertainty estimates does not automatically produce well-calibrated uncertainty—predictive intervals that correctly capture the stated confidence level in practice—and rigorous calibration assessment against held-out or synthetic ground truth remains an essential but frequently underemphasized component of AI-driven UQ development for GCS.

Scaling to three-dimensional, century-scale problems. While dimension-adaptive architectures (Mahjour et al., 2025) represent meaningful progress, fully three-dimensional, coupled flow-geomechanics-geochemistry UQ at the century-scale time horizons relevant to long-term storage security remains computationally demanding, and further architectural and algorithmic innovation will be required to make such comprehensive UQ routine at commercial scale.

Validation over unobservable time horizons. Because direct empirical validation of century-scale containment predictions is fundamentally infeasible within any practical research timeframe, AI-driven UQ frameworks for long-term storage security must rely on in silico validation strategies (Gahlot et al., 2025) whose own limitations—dependence on the fidelity of the underlying simulation models used to generate synthetic validation data—require continued scrutiny and methodological development.

Translating uncertainty into regulatory decision criteria. Even well-calibrated uncertainty estimates require translation into specific, actionable regulatory decision criteria—acceptable probability thresholds for containment failure, appropriate monitoring frequency given quantified uncertainty levels—and the standardization of these decision criteria across regulatory jurisdictions remains an ongoing institutional challenge distinct from, but essential to, the underlying technical UQ methodology.

Integration across coupled physical processes. Much current AI-driven UQ research addresses flow and pressure uncertainty in relative isolation from geomechanical and geochemical uncertainty, and comprehensive, jointly propagated uncertainty across the full range of coupled physical processes relevant to long-term storage security remains an important, only partially realized research frontier.

8. Future Directions

Future research should prioritize the extension of dimension-adaptive Bayesian neural network approaches (Mahjour et al., 2025) and deep convolutional encoder-decoder surrogates (Mo et al., 2019) toward fully coupled flow-geomechanics-geochemistry uncertainty propagation, enabling comprehensive, jointly quantified storage security assessment rather than uncertainty characterization of isolated physical processes. Closer integration of AI-enhanced data assimilation frameworks (Seabra et al., 2024) with uncertainty-aware digital shadow monitoring architectures (Gahlot et al., 2025) represents a promising direction for developing continuously updated, long-term security assessment systems capable of progressively sharpening epistemic uncertainty as monitoring data accumulates over the multi-decade project lifetime.

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 AI-driven UQ training pipelines represents a further avenue for improving the fidelity of site-specific uncertainty estimates, with particular relevance to heterogeneous, multi-constraint storage sites such as the Evetts site in the Permian Basin. Finally, adopting communication, governance, and cost-justification frameworks informed by uncertainty-aware AI deployment in national-scale healthcare systems (Nguyen, 2026a, 2026c, 2026d) may help establish the standardized decision criteria and stakeholder communication protocols required for AI-driven UQ to transition from research demonstration toward routine, regulator-accepted use in long-term commercial-scale CO_2 storage security assessment.

9. Conclusion

Artificial intelligence-driven uncertainty quantification is an essential, and increasingly central, complement to point-prediction machine learning for ensuring the long-term security of geological CO_2 storage. Applications spanning dimension-adaptive Bayesian neural networks, deep convolutional encoder-decoder surrogates, AI-enhanced data assimilation, and uncertainty-aware digital shadow monitoring collectively demonstrate that explicitly characterizing and propagating both aleatoric and epistemic uncertainty is achievable at computational costs substantially lower than conventional Monte Carlo approaches, while producing the calibrated, probabilistic containment metrics that regulatory decision-making increasingly requires. Cross-domain lessons from uncertainty-aware AI deployment in national-scale healthcare systems further suggest a pathway toward the governance, communication, and cost-justification frameworks necessary for AI-driven UQ to achieve regulatory trust and routine operational adoption. Realizing the full potential of AI-driven UQ for long-term CO_2 storage security will require sustained investment in uncertainty calibration, computational scalability to fully coupled, three-dimensional, century-scale problems, and the translation of quantified uncertainty into standardized regulatory decision criteria—investments essential to ensuring that geological CO_2 storage can be trusted, monitored, and managed responsibly over the multi-century timescales required to deliver genuine climate mitigation value.

References

1. Gahlot, A. P., Orozco, R., Yin, Z., & Herrmann, F. J. (2025). An uncertainty-aware digital shadow for underground multimodal CO2 storage monitoring. Geophysical Journal International, 242(1), ggaf176. https://doi.org/10.1093/gji/ggaf176
2. Mahjour, S. K., Saleh, A., & Mahjour, S. S. (2025). Dimension-adaptive machine learning for efficient uncertainty quantification in geological carbon storage models. Processes, 13(6), 1834. https://doi.org/10.3390/pr13061834
3. Mo, S., Zhu, Y., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. Water Resources Research, 55(1), 703–728. https://doi.org/10.1029/2018WR023528
4. 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.
5. 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.
6. Nguyen, T. T. (2026d). National-scale predictive analytics for Medicare and Medicaid: An AI-driven approach to identifying high-risk populations and reducing healthcare costs. Journal of Advanced Scientific Research, 17(6), 1–10.
7. Seabra, G. S., Mücke, N. T., Silva, V. L. S., Voskov, D., & Vossepoel, F. C. (2024). AI enhanced data assimilation and uncertainty quantification applied to geological carbon storage. International Journal of Greenhouse Gas Control, 136, 104190. https://doi.org/10.1016/j.ijggc.2024.104190.